

UNITED STATES AIR FORCE
SUMMER RESEARCH PROGRAM -- 1997
SUMMER RESEARCH EXTENSION PROGRAM FINAL REPORTS

VOLUME 4A
WRIGHT LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES
5800 Uplander Way
Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Major Linda Steel-Goodwin

Program Manager, RDL
Scott Licoscas

Program Administrator, RDL
Johnetta Thompson

Program Administrator, RDL
Rebecca Kelly-Clemmons

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
Bolling Air Force Base
Washington, D.C.
December 1997

20010319 009

AQMO1-06-1189

PREFACE

This volume is part of a five-volume set that summarizes the research of participants in the 1997 AFOSR Summer Research Extension Program (SREP.)

Reports presented in this volume are arranged alphabetically by author and are numbered consecutively – e.g., 1-1, 1-2, 1-3; 2-1, 2-2, 2-3. Reports in the five-volume set are organized as follows:

VOLUME

TITLE

1	Armstrong Laboratory
2	Phillips Laboratory
3	Rome Laboratory
4A	Wright Laboratory
4B	Wright Laboratory
5	Arnold Engineering Development Center Air Logistics Centers United States Air Force Academy Wilford Hall Medical Center

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering the data, reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paper Project Collection (0704-0188), Washington, DC 20503.

AFRL-SR-BL-TR-00-

reviewing
information

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE December, 1997		3. REPORT NUMBER 0711	
4. TITLE AND SUBTITLE 1997 Summer Research Program (SRP), Summer Research Extension Program (SREP), Final Report, Volume 4A, Wright Laboratory				5. FUNDING NUMBERS F49620-93-C-0063	
6. AUTHOR(S) Gary Moore					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Research & Development Laboratories (RDL) 5800 Uplander Way Culver City, CA 90230-6608				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research (AFOSR) 801 N. Randolph St. Arlington, VA 22203-1977				10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES					
12a. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release				12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) The United States Air Force Summer Research Program (SRP) is designed to introduce university, college, and technical institute faculty members to Air Force research. This is accomplished by the faculty members, graduate students, and high school students being selected on a nationally advertised competitive basis during the summer intersession period to perform research at Air Force Research Laboratory (AFRL) Technical Directorates and Air Force Air Logistics Centers (ALC). AFOSR also offers its research associates (faculty only) an opportunity, under the Summer Research Extension Program (SREP), to continue their AFOSR-sponsored research at their home institutions through the award of research grants. This volume consists of a listing of the participants for the SREP and the technical report from each participant working at the AF Wright Laboratory.					
14. SUBJECT TERMS Air Force Research, Air Force, Engineering, Laboratories, Reports, Summer, Universities, Faculty, Graduate Student, High School Student				15. NUMBER OF PAGES	
				16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified		18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified		19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	
				20. LIMITATION OF ABSTRACT UL	

1997 SREP Final Technical Report Table of Contents

Armstrong Laboratory

Volume 1

	Principle Investigator	Report Title University/Institution	Laboratory & Directorate
1	Dr. Richelle M. Allen-King	Trans-1,2-Dichloroethene Transformation Rate in a Metallic Iron/Water System: Effects of Concentration and Temperature Washington State University	AL/EQC
2	Dr. Anthony R. Andrews	Development of Multianalyte Electrochemiluminescence Sensors & Biosensors Ohio University	AL/EQC
3	Dr. Jer-Sen Chen	Development of Perception Based Video Compression Algorithms Using Reconfigurable Hardware Wright State University	AL/CFHV
4	Dr. Cheng Cheng	Investigation & Eval of Optimization Algorithms Guiding the Assignment of Recruits to Training School Seats John Hopkins University	AL/HRM
5	Dr. Randolph D. Glickman	Optical Detection of Intracellular Photooxidative Reactions University of Texas Health Science Center	AL/OEO
6	Dr. Nandini Kannan	Predicting Altitude Decompression sickness Using Survival Models University of Texas at San Antonio	AL/CFTS
7	Dr. Antti J. Koivo	Skill Improvements Via Reflected Force Feedback Purdue Research Foundation	AL/CFBA
8	Dr. Suk B. Kong	Degradation & Toxicology Studies of JP-8 Fuel in Air, Soil & Drinking Water Incarnate Word College	AL/OEA
9	Dr. Audrey D. Levine	Biogeochemical Assessment of Natl Attenuation of JP-4 Contaminated Ground in the Presence of Fluorinated Surfactants Utah State University	AL/EQC
10	Dr. Robert G. Main	The Effect of Video Image Size & Screen Refresher Rate On Mess Retention Cal State University, Chico	AL/HRT
11	Dr. Phillip H. Marshall	On the Resilience of Time-to-Contact Judgements: The Determination of Inhibitory and Facilitory Influences, and Factor Structure Texas Tech University	AL/HRM
12	Dr. Bruce V. Mutter	Environmental cost Analysis: Calculating Return on Investment for Emerging Technologies Bluefield State College	AL/EQP

1997 SREP Final Technical Report Table of Contents

Armstrong Laboratory

Volume 1 (cont.)

	Principle Investigator	Report Title University/Institution	Laboratory & Directorate
13	Dr. Sundaram Narayanan	Java-Based Interactive Simulation Architecture for Airbase Logistics Modeling Wright State University	AL/HRT
14	Dr. Barth F. Smets	Coupling of 2, 4-&2, 6-Dinitrotoluene Mineralization W/NO2 Removal by University of Cincinnati	AL/EQC
15	Dr. Mary Alice Smith	In Vitro Detection of Apoptosis in Differentiating Mesenchymal Cells Using Immunohistochemistry and Image Analysis University of Georgia	AL/OET
16	Dr. William A. Stock	Application of Meta-Analysis to Research on Pilot Training: Extensions to Flight Simulator Visual System Research Arizona State University	AL/HRA
17	Dr. Nancy J. Stone	Evaluation of a Scale Designed to Measure the Underlying Constructs of Engagement, Involvement, & Self-Regulated Learning Creighton University	AL/HRT
18	Dr. Mariusz Ziejewski	Characterization of Human Head/Neck Response in Z-Direction in Terms of Significant Anthropomorphic Parameters, Gender, Helmet Weight and Helmet Center North Dakota State University	AL/CFBV
19	Dr. Kevin M. Lambert	Magnetic Effects on the Deposition & Dissolution of Calcium Carbonate Scale Brigham Young University	AL/EQS
20	Dr. Jacqueline C. Shin	Coordination of Cognitive & Perceptual-Motor Activities Pennsylvania State University	AL/HRM
21	Dr. Travis C. Tubre	The Development of a General Measure of Performance Texas A&M University-College Station	AL/HRT
22	Dr. Robert B. Trelease	Development of Qualitative Process Modeling Systems for Cytokines, Cell Adhesion Molecules, and Gene Regulation University of California – Los Angeles	AL/AOH

1997 SREP Final Technical Report Table of Contents

Phillips Laboratory

Volume 2

	Principle Investigator	Report Title University/Institution	Laboratory & Directorate
1	Dr. Graham R. Allan	Temporal & Spatial Characterization of a Synchronously-Pumped New Mexico Highlands University	PL/LIDN
2	Dr. Joseph M. Calo	Transient Studies of the Effects of Fire Suppressants in a Well-Stirred Combustor Brown University	PL/GPID
3	Dr. James J. Carroll	Examination of Critical Issues in the Triggering of Gamma Rays from 178Hfm2 Youngstown State University	PL/WSQ
4	Dr. Soyoung S. Cha	Gradient-Data Tomography for Hartman Sensor Application to Aero-Optical Field Reconstruction University of Illinois at Chicago	PL/LIMS
5	Dr. Judith E. Dayhoff	Dynamic Neural Networks: Towards Control of Optical Air Flow Distortions University of Maryland	PL/LIMS
6	Dr. Ronald R. DeLyser	Computational Evaluation of Optical Sensors University of Denver	PL/WSTS
7	Dr. Andrew G. Detwiler	Analysis & Interpretation of Contrail Formation Theory & Observations South Dakota School of Mines – Technology	PL/GPAB
8	Dr. Itzhak Dotan	Measurements of Ion-Molecule Reactions at Very High Temperature The Open University of Israel	PL/GPID
9	Dr. George W. Hanson	Electromagnetic Modeling of Complex Dielectric/Metallic Mines In A Layered University of Wisconsin – Milwaukee	PL/WSQ
10	Dr. Mayer Humi	Optical & Clear Air Turbulence Worcester Polytechnic Inst.	PL/GPAA
11	Dr. Christopher H. Jenkins	Shape Control of an Inflated Circular Disk Experimental Investigation South Dakota School of Mines – Technology	PL/VT
12	Dr. Dikshitulu K. Kalluri	Numerical Simulation of Electromagnetic Wave Transformation in a Dynamic Magnetized Plasma University of Lowell	PL/GPIA
13	Dr. Aravinda Kar	Improved Chemical Oxygen-Iodine Laser (COIL) Cutting Models to Optimize Laser Parameters University of Central Florida	PL/LIDB

1997 SREP Final Technical Report Table of Contents

Phillips Laboratory

Volume 2 (cont.)

	Principle Investigator	Report Title University/Institution	Laboratory & Directorate
14	Dr. Andre Y. Lee	Characterization of Thermoplastic Inorganic-Organic Hybrid Polymers Michigan State University	PL/RKS
15	Dr. Feng-Bao Lin	Improvement in Fracture Propagation Modeling for Structural Ballistic Risk Assessment Polytechnic University of New York	PL/RKEM
16	Dr. Ronald A. Madler	Cross Sectional Area Estimation of Orbital Debris Embry-Riddle Aeronautical University	PL/WSAT
17	Dr. Carlos A. Ordonez	Incorporation of Boundary condition Models into the AF Computer Simulation University of North Texas	PL/WSQA
18	Dr. James M. Stiles	Wide Swath, High Resolution, Low Ambiguity SAR Using Digital Beamforming Arrays University of Kansas	PL/VTRA
19	Dr. Charles M. Swenson	Balloon Retromodulator Experiment Post- flight Analysis Utah State University	PL/VTRA
20	Dr. Miguel Velez-Reyes	Development of Algorithms for Linear & Nonlinear Retrieval Problems in Atmospheric Remote Sensing University of Puerto Rico	PL/GPAS
21	Dr. John D. Holtzclaw	Experimental Investigation of Ipinging Jets University of Cincinnati	PL/RKS
22	Dr. Jeffrey W. Nicholson	Radar Waves with Optical Carriers University of New Mexico	PL/LIDB

1997 SREP Final Technical Report Table of Contents

Rome Laboratory

Volume 3

Principle Investigator	Report Title University/Institution	Laboratory & Directorate
1 Dr. A. F. Anwar	Deep Quantum Well Channels for Ultra Low Noise HEMTs for Millimeter and Sub-millimeter Wave Applications University of Connecticut	AFRL/SNH
2 Dr. Ahmed E. Barbour	Investigating the Algorithmic Nature of the Proof Structure of ORA Larch/VHDL Georgia Southern University	RL/ERDD
3 Dr. Milica Barjaktarovic	Specification & Verification of MISSI Architecture Using SPIN Wilkes University	RL/C3AB
4 Dr. Daniel C. Bukofzer	Analysis, Performance Evaluations, & Computer Simulations of Receivers Processing Low Probability of Intercept Signals Cal State Univ. Fresno	RL/C3BA
5 Dr. Xuesheng Chen	Non-Destructive Optical Characterization of Composition & Its Uniformity in Multilayer Ternary Semiconductor Stacks Wheaton College	RL/ERX
6 Dr. Jun Chen	Amplitude Modulation Using Feedback Sustained Pulsation as Sub-Carrier in Rochester Inst of Technol	RL/OCPA
7 Dr. Everett E. Crisman	Development of Anti-Reflection Thin Films for Improved Coupling of Laser Energy into Light Activated, Semiconductor Re-Configurable, Microwave Source/Antenna Brown University	RL/ERAC
8 Dr. Digendra K. Das	Development of a Stimulation Model for Determining the Precision Of Reliability SUNYIT	RL/ERSR
9 Dr. Matthew E. Edwards	An Application of PROFILER for Modeling the Diffusion of Of Aluminum-Copper on a Silicon Substrate Spelman College	RL/ERDR
10 Dr. Kaliappan Gopalan	Analysis of Stressed Speech Using Cepstral Domain Features Purdue University - Calumet	RL/IRAA
11 Dr. James P. LeBlanc	Multichannel Autoregressive Modeling & Multichannel Innovations Based New Mexico State University	RL/OCSS
12 Dr. Hrushikesh N. Mhaskar	Multi-Source Direction Finding Cal State University, Los Angeles	RL/ERAA
13 Dr. Ronald W. Noel	An Evolutionary Sys for Machine Recognition of Software Source Code Rensselaer Polytechnic Inst	RL/C3CA

1997 SREP Final Technical Report Table of Contents

Rome Laboratory

Volume 3 (cont.)

Principle Investigator	Report Title University/Institution	Laboratory & Directorate
14 Dr. Glenn E. Prescott	Rapid Prototyping of Software Radio Sys Using Field Programmable Gate Arrays University of Kansas Center for Research	RL/C3BB
15 Dr. Mysore R. Rao	Wavelet Transform Techniques for Isolation, Detection & Classification of Concealed Objects in Images Rochester Institute of Technology	RL/OCSM
16 Dr. Scott E. Spetka	IPL HTML Interface Performance Evaluation SUNY of Tech Utica	RL/IRD
17 Dr. Gang Sun	Investigation of Si/ZnS Near Infrared Intersubband Lasers University of Massachusetts-Boston	RL/EROOC
18 Mr. Parker E. Bradley	Development of a User-Friendly Computer Environment for Blind Source Syracuse University	RL/C3BB

1997 SREP Final Technical Report Table of Contents

Wright Laboratory

Volume 4A

Principle Investigator	Report Title University/Institution	Laboratory & Directorate
1 Dr. Mohammad S. Alam	Infrared Image Registration & High Resolution Reconstruction Using Rotationally Translated Video Sequences* Purdue University	WL/AAJT
2 Dr. Prina Ari-Gur	Optimizing Microstructure, Texture & Orientation Image Microscopy of Hot Rolled Ti-6Al-4V Western Michigan University	WL/MLLN
3 Dr. James D. Baldwin	Multi-Site & Widespread Fatigue Damage in Aircraft Structure in the Presence of Prior Corrosion University of Oklahoma	WL/FIB
4 Dr. Armando R. Barreto	Deconvolution of the Space-Time Radar Spectrum Florida International University	WL/AAMR
5 Dr. Marc M. Cahay	Improved Modeling of Space-Charge Effects in a New Cold Cathode Emitter University of Cincinnati	WL/AADM
6 Dr. Reaz A. Chaudhuri	Interfacing of Local Asymptotic Singular & Global Axisymmetric Micromechanical University of Utah	WL/MLBM
7 Dr. Robert J. DeAngelis	Texture Formation During the Thermo-Mechanical Processing of Copper Plate University of Nebraska – Lincoln	WL/MNMW
8 Dr. Gregory S. Elliott	The Study of a Transverse Jet in a Supersonic Cross-Flow Using Advanced Laser Rutgers: State University of New Jersey	WL/POPT
9 Dr. Altan M. Ferendeci	Development of Multiple Metal-Dielectric Layers for 3-D MMIC University of Cincinnati	WL/AADI
10 Dr. Allen G. Greenwood	Development of a Prototype to Test & Demonstrate the MODDCE Framework Mississippi State University	WL/MTI
11 Dr. Michael A. Grinfeld	Mismatch Stresses & Lamellar Microstructure of TiAl-Alloys Rutgers University- Piscataway	WL/MLLM
12 Dr. Michael C. Larson	Interfacial Sliding in Brittle Fibrous Composites Tulane University	WL/MLLM
13 Dr. Douglas A. Lawrence	Tools for the Analysis & Design of Gain Scheduled Missile Autopilots Ohio University	WL/MNAG

1997 SREP Final Technical Report Table of Contents

Wright Laboratory (cont.)

Volume 4A

	Principle Investigator	Report Title University/Institution	Laboratory & Directorate
14	Dr. Junghsen Lieh	Determination of 3D Deformations, Forces & Moments of Aircraft Wright State University	WL/FIVM
15	Dr. Zongli Lin	Control of Linear Sys w/Rate Limited Actuators & Its Applications to Flight Control Systems SUNY Stony Brook	WL/FI
16	Dr. Paul Marshall	Experimental & Computational Investigations of Bromine & Iodine Chemistry in Flame Suppression University of North Texas	WL/MLBT
17	Dr. Hui Meng	Development of Holographic Visualization & Holographic Velocimetry Techniques Kansas State University	WL/POSC
18	Dr. Douglas J. Miller	Band Gap Calculations on Squarate-Containing Conjugated Oligomers for the Prediction of Conductive and Non-Linear Optical Properties of Polymeric Materials Cedarville College	WL/MLBP
19	Dr. Timothy S. Newman	Classification & Visualization of Tissue in Multiple Modalities of Brain MR University of Alabama at Huntsville	WL/AACR
20	Dr. Mohammed Y. Niamat	FPGA Implementation of the Xpatch Ray Tracer University of Toledo	WL/AAST
21	Dr. Anthony C. Okafor	Development of Optimum Drilling Process for Advanced Composites University of Missouri – Rolla	WL/MTI
22	Dr. George A. Petersson	Absolute Rates for Chemical Reactions Wesleyan University	WL/MLBT
23	Dr. Mohamed N. Rahaman	Process Modeling of the Densification of Granular Ceramics Interaction Between Densification and Creep University of Missouri – Rolla	WL/MLLN

1997 SREP Final Technical Report Table of Contents

Wright Laboratory (cont.)

Volume 4B

	Principle Investigator	Report Title University/Institution	Laboratory & Directorate
24	Dr. Martin Schwartz	Quantum Mechanical Modeling of the Thermochemistry of Halogenated Fire Suppressants University of North Texas	WL/MLBT
25	Dr. Marek Skowronski	Investigation of Slip Boundaries in 4H-SiC Crystals Carnegie Melon University	WL/MLPO
26	Dr. Yong D. Song	Guidance & Control of Missile Sys Under Uncertain Flight Conditions North Carolina A&T State University	WL/MNAG
27	Dr. Raghavan Srinivasan	Models for Microstructural Evolution During Dynamic Recovery Wright State University	WL/MLIM
28	Dr. Scott K. Thomas	The Effects of Transient Acceleration Loadings on the Performance of a Copper-Ethanol Heat Pipe with Spiral Grooves Wright State University	WL/POOS
29	Dr. James P. Thomas	The Effect of Temperature on Fatigue Crack Growth of TI-6AL-4V in the Ripple University of Notre Dame	WL/MLLN
30	Dr. Karen A. Tomko	Scalable Parallel Solution of the 3D Navier-Stokes Equations Wright State University	WL/FIM
31	Dr. J. M. Wolff	Off Design Inviscid/Viscous Forced Response Prediction Model for High Cycle Wright State University	WL/POTF
32	Mr. Todd C. Hathaway	Experiments on Consolidation of Aluminum Powders Using Simple Shear University of North Texas	WL/MLLN
33	Ms. Diana M. Hayes	Error Correction & Compensation for Mueller Matrices Accounting for Imperfect Polarizers University of North Texas	WL/MNGA

1997 SREP Final Technical Report Table of Contents

Volume 5

	Principle Investigator	Report Title University/Institution	Laboratory & Directorate
		Arnold Engineering Development Center	
1	Dr. Frank G. Collins	Development of Laser Vapor Screen Flow Visualization Sys Tennessee University Space Institute	AEDC
		United States Air Force Academy	
2	Mr. Derek E. Lang	Experimental Investigation of Liquid Crystal Applications for Boundary Layer Characterization University of Washington	USAFA/DFA
		Air Logistics Centers	
3	Dr. Sandra A. Ashford	Development of Jet Engine Test Facility Vibration Signature & Diagnostic System University of Detroit Mercy	OCALC/TIE
4	Dr. Roger G. Ford	Use of Statistical Process Control in a Repair/Refurbish/ Remanufactureg Environment St. Mary's University	SAALC
		Wilford Hall Medical Center	
5	Dr. Stedra L. Stillmana	Metabolite Profile Following the Administration of Fenproporex University of Alabama at Birmingham	WHMC

INFRARED IMAGE REGISTRATION AND HIGH RESOLUTION RECONSTRUCTION USING ROTATIONALLY TRANSLATED VIDEO SEQUENCES*

Mohammad S. Alam

Associate Professor of Electrical Engineering

Purdue University, ET 327A, Fort Wayne, IN 46805-1499

Voice: 219-481-6020, Email: alam@engr.ipfw.edu

Final Project Report for:

Summer Research Extension Program

Wright Laboratory, Wright-Patterson Air Force Base, Dayton, Ohio

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, Washington DC

and

Wright Laboratory, Wright-Patterson Air Force Base

December 1997

Infrared Image Registration And High Resolution Reconstruction Using Rotationally Translated Video Sequences*

Mohammad S. Alam

Associate Professor of Electrical Engineering

Purdue University, ET 327A, Fort Wayne, IN 46805-1499

Abstract

Microscanning is an effective technique for reducing aliasing and increasing resolution in images produced by staring infrared imaging systems, which involves recording a sequence of frames through subpixel movements of the field of view on the detector array and then interlacing them to produce a high resolution image. The forward looking infrared (FLIR) detector arrays is usually mounted on a moving platform, such as an aircraft, and the normal vibrations associated with the moving platform can be used to generate shifts in the FLIR recorded sequence of frames. Among the various image registration algorithms, the algorithm proposed by Irani and Peleg¹ has shown the potential of being an extremely fast and efficient method of registering a sequence of image frames. This algorithm has been demonstrated to work with non uniformity-corrected, aliased imagery collected from a 3 to 5 micrometer thermal imaging system. However, since the algorithm is gradient-based and depends on its ability to accurately calculate spatial changes between images, its performance over all imaging conditions is unknown. For example, it is expected that fixed-pattern spatial noise will produce errors in the gradient estimate. Other factors that may affect the accuracy of registration include the severity of aliasing in the imagery, image contrast, and image size. This research project was directed to investigate the effect of the aforementioned parameters on the image registration algorithm.

* When this project was funded in March 1997, the Air Force Lab Focal Point (Brian Yasuda, WL/AAJT) informed the PI that another group at the Wright Laboratory completed the investigation of high resolution image reconstruction using rotationally translated frames and they are more interested in investigating the effects of various degrading factor on the registration algorithm proposed by Irani and Peleg¹. After detailed discussion with the focal point and other group members, it was decided to investigate the performance of the registration algorithm under various degrading factors such as noise, frame size, aliasing and image contrast.

Infrared Image Registration And High Resolution Reconstruction Using Rotationally Translated Video Sequences *

Mohammad S. Alam

Associate Professor of Electrical Engineering

Purdue University, ET 327A, Fort Wayne, IN 46805-1499

1 Introduction

In many infrared imaging systems, the detector spacing in the focal plane array is not sufficiently small so as to sample a band-limited scene at the Nyquist rate, resulting in a degraded image due to aliasing artifacts. Over the last five years, the Electro-Optics Branch of Wright Laboratory has been developing a microscan imaging technique to increase the spatial sampling rate of existing focal plane arrays and to alleviate the aliasing effects in infrared imagery^{2,3}. The microscanning process uses a sequence of spatially undersampled time frames of a scene to generate the high resolution image. Each frame is subpixel shifted relative to each other frame onto a set grid pattern. The sequence of frames are then interlaced to yield the high resolution image which represents the input scene effectively sampled at a higher spatial frequency. The aforementioned process is called controlled microscanning because the subpixel shifts between the temporal image frames are controlled and therefore known *a priori*. Uncontrolled microscanning is the process where the shifts for each frame are unknown and must be estimated before registering onto the high resolution grid pattern. The shifts for each frame are unknown because they are generated by the random motion and/or vibration of the imager platform such as an aircraft and not by a microscan mirror.

Among the various image registration algorithms proposed in the literature, the image registration algorithm reported in Reference 1 has been found to be particularly attractive. This registration algorithm has shown the potential of being an extremely fast and efficient method of registering a sequence of image frames. The algorithm has been demonstrated to work with non uniformity-corrected, aliased imagery collected from a laboratory 3- to 5- μm thermal imaging

system. However, since the algorithm is gradient-based and depends on its ability to accurately calculate spatial changes between images, its performance over all imaging conditions is unknown. For example, it is expected that fixed-pattern spatial noise will produce errors in the gradient estimate. Other factors that may affect the accuracy of registration include the severity of aliasing in the imagery, image contrast, and image size. Therefore, an evaluation of the performance of the registration algorithm proposed in Reference 1 may involve determining a method to characterize the algorithm's performance under various levels of degradation. The main degrading factors to be examined include fixed-pattern spatial noise, temporal noise, aliasing, contrast, and image size. A measure of the severity of the degradation under consideration must be determined. Plots of registration error versus degradation measure may be a useful indicator of general trends. It may also be informative to combine the degrading factors in some manner to determine the total effects on registration error.

To investigate the effect of various degrading factors on above mentioned registration algorithm¹, a 450×450 pixel infrared image data collected on a real time infrared imaging system was used as the high resolution image. Since all parameters of the imager optics and detector array are known, it is possible to accurately model the optical transfer function and the detector transfer function which are then used to obtain the system's modulation transfer function (MTF). The high resolution image was blurred by the system's MTF and a sequence of nine low resolution frames were generated from this blurred image.

Various amounts of aliasing and noise was added to the aforementioned image data which are then used as input to the registration algorithm. In addition, a number of various contrast imagery has been produced from simulations of the identical image to correspond to different times of day and year. The contrast of the imagery was adjusted to represent the contrast in typical infrared imagery as closely as possible. From the simulation results, it is evident that except for the spatio-temporal noise, the registration algorithm is susceptible to various levels of registration error with respect to the other degrading parameters such as the fixed pattern noise, aliasing, image size and image contrast.

2 The Sampling Process

2.1 Sampling in a FLIR Array

In a FLIR array, sampling is performed by a finite sized array of detectors and three main factors must be taken into consideration: the optical point spread function, the detector charge integration, and the detector array geometry. The block diagram in Fig. 1 illustrates the sampling process in a FLIR array. The object scene, denoted $o(x,y)$, is convolved with the point spread function of the optics, $psf(x,y)$, and the aperture function for the square detector³, $d(x,y)$.

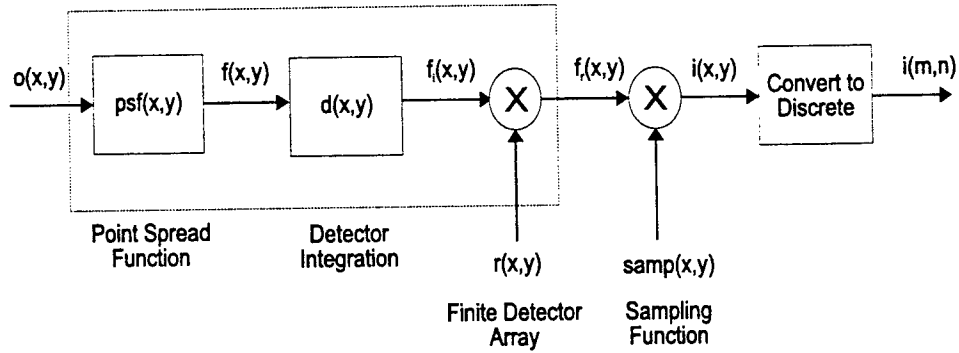


Figure 1. Block diagram of the imaging system.

For our imaging system, we assume the detector has a flat response across its active region so the detector function can be expressed as:

$$d(x,y) = \frac{1}{|ab|} \text{rect}\left(\frac{x}{a}, \frac{y}{b}\right), \quad (1)$$

where a and b are the dimensions of the individual detectors. The result is multiplied by a function representing the limited extent of the detector array, $r(x,y)$, expressed as

$$r(x,y) = \text{rect}\left(\frac{x}{X}, \frac{y}{Y}\right), \quad (2)$$

where X and Y are the dimensions of the array. To apply this integration to all of the detectors, it is multiplied by the sampling lattice, $\text{comb}_{\Delta x, \Delta y}(x, y)$,

$$\text{comb}_{\Delta x, \Delta y}(x, y) = \frac{1}{\Delta x \Delta y} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta\left(\frac{x}{\Delta x} - m, \frac{y}{\Delta y} - n\right) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta(x - m\Delta x, y - n\Delta y) \quad (3)$$

where Δx and Δy are the center-to-center detector spacing. The resulting expression for the staring image is

$$i(x, y) = \left[o(x, y) * \text{psf}(x, y) * \frac{1}{|ab|} \text{rect}\left(\frac{x}{a}, \frac{y}{b}\right) \right] \cdot \text{rect}\left(\frac{x}{X}, \frac{y}{Y}\right) \cdot \text{comb}_{\Delta x, \Delta y}(x, y). \quad (4)$$

An illustration of a uniform detector array showing critical dimensions is shown in Fig. 2.

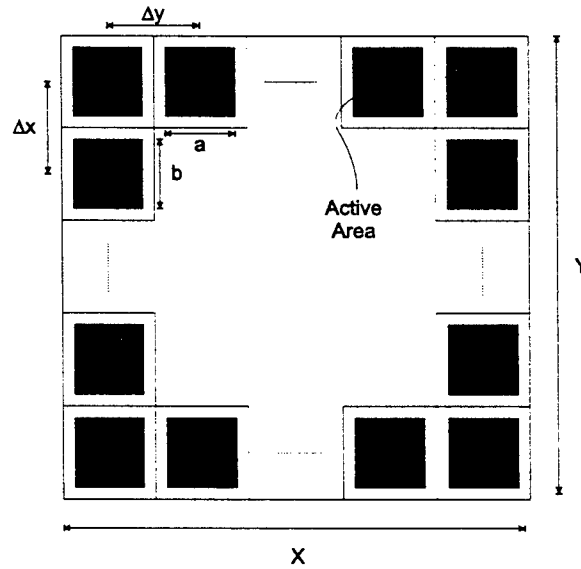


Figure 2. The detector array.

2.2 Controlled Microscanning

In controlled microscanning, the subpixel shifts between temporal image frames are controlled and therefore known in advance. A level L microscan is defined to be the case where L^2 staring

frames are acquired for each high resolution frame and each staring frame has its own unique controlled subpixel shift, each shift being part of a uniform grid. Thus, each staring frame has a shift that is integer multiple of $1/L$ times the detector width. The staring frames are interlaced in an $L \times L$ pattern to produce a high resolution frame of size $NL \times NL$ where N is the size of the square detector array. For a level 2 or 2×2 microscanned image, the original scene is stepped one half the length of the detector in the x and y directions producing a series of four staring images. These images are then interlaced to produce the resulting microscan image of $2N \times 2N$ pixels, as shown in Fig. 3 where the reconstructed image has 4 times the number of unique samples as any of the four frames. Thus, a level- L microscanning effectively increases the sampling frequency by L without changing the detector size or spacing. Since the effect of the finite detector array $r(x,y)$ is small compared to the effects of the detector integration and the psf, neglecting the effect of $r(x,y)$, an $L \times L$ microscan process can be expressed as

$$i(x, y) = \left[o(x, y) * psf(x, y) * \frac{1}{|ab|} rect\left(\frac{x}{a}, \frac{y}{b}\right) \right] \frac{1}{L^2} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} comb_{\Delta x, \Delta y} \left(x - \frac{i}{L}, y - \frac{j}{L} \right) \quad (5)$$

where the factor of $1/L^2$ is used to adjust for the reduction in detector integration time at each of the L^2 microscan steps.

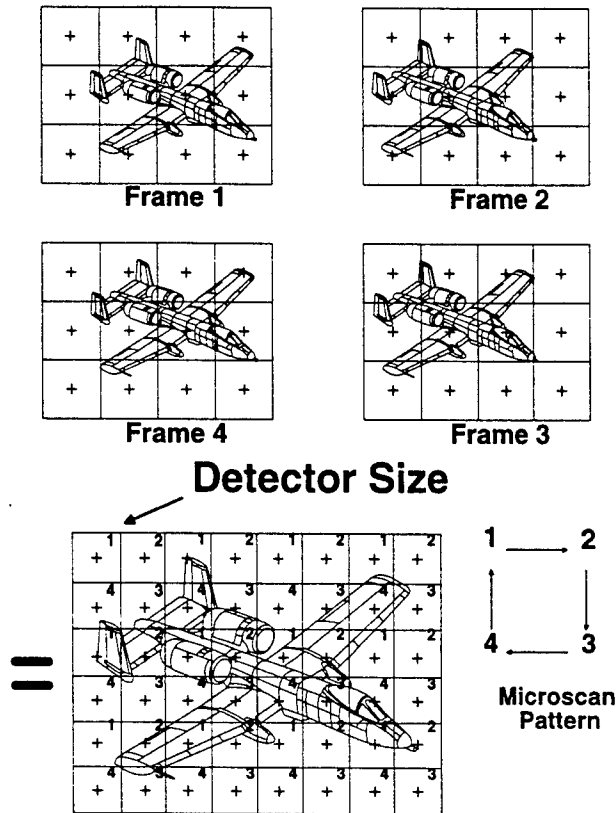


Figure 3. Illustration of a level 2 microscanning process.

2.3 Uncontrolled Microscanning

Uncontrolled microscanning results from practical applications where the imager is mounted on a moving and/or vibrating platform, such as an aircraft, and the vibrations associated with the platform can be exploited to create the shifts in the acquired frames. In uncontrolled microscanning, the shifts for each recorded frame are unknown and must be obtained before forming an estimate of the high resolution image. Because the shifts are random in uncontrolled microscanning, it eliminates the need for a microscan mirror and driver system otherwise required in controlled microscanning. Since a fixed number of frames is required for a given level of microscanning, and the shifts are random, some of the acquired frames may have almost similar shifts or unusable shifts resulting in one or more missing frames. The missing frames will generate empty bins in the high resolution image reconstruction. Thus, the key factors for the high resolution image reconstruction are the accurate knowledge of the subpixel translational motion of the scene relative to the FLIR array and the accurate estimation of the missing frames

for filling the empty bins in the high resolution image. These missing frames may be estimated using various algorithms such as the maximum likelihood algorithm⁴, maximum *a posteriori* algorithm⁵ or the weighted nearest neighbor algorithm⁶ depending on trade-offs such as the processing speed.

3 Image Registration

For uncontrolled microscanning, the movement of the scene on the array is produced by random motion and/or vibration of the imager platform. Therefore the image shifts are unknown for each recorded frame. It is necessary to develop a suitable algorithm which can accurately estimate the registration parameters i.e., in our case, the random translational shifts. Various algorithms have been reported in the literature for estimating the image registration parameters^{1,7,8}. Among these techniques, the gradient based registration technique proposed in Ref. 7 appears to be particularly attractive for practical applications. However, this technique only works for small shifts. To accommodate the case where larger shift values are anticipated, we utilized the iterative technique proposed in Reference 1 and incorporated it with the registration algorithm proposed in Reference 7. In this algorithm, the first acquired frame is considered to be the reference frame and the shifts associated with the remaining frames are calculated with respect to the reference frame. If p represents the total number of frames acquired by the imager, and h_k and v_k represents the shifts in the horizontal and vertical directions for the k th frame, then the observed k th frame may be expressed as

$$o_k(x, y) = o_1(x + h_k, y + v_k), \quad (6)$$

where $k \in \{2, 3, 4, \dots, p\}$ and for the reference frame, $h_1 = v_1 = 0$. Considering the first three terms for the Taylor series expansion, Eq. (6) may be approximated as

$$o_k(x, y) \approx o_1(x, y) + h_k \frac{\partial o_1(x, y)}{\partial x} + v_k \frac{\partial o_1(x, y)}{\partial y}. \quad (7)$$

Since Eq. (7) is an approximation itself and discrete estimates of x and y must be used, it is useful to apply the method of least squares for solving the registration parameters h_k and v_k . For the least squares solution⁹ the error expression

$$E_k(h_k, v_k) \approx \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N \left[o_k(m, n) - o_1(m, n) - h_k \frac{\partial o_1(m, n)}{\partial m} - v_k \frac{\partial o_1(m, n)}{\partial n} \right]^2 \quad (8)$$

must be minimized, where m and n are discrete variables for the x and y directions, M and N represents the total number of pixels in the x and y directions, and h_k and v_k are the translational shifts in the x and y directions between the k th frame and reference frame. Equation (8) can be minimized by differentiating $E_k(h_k, v_k)$ with respect to h_k and v_k and setting the derivatives equal to zero. This yields two equations which must be solved simultaneously and can be conveniently represented in the following matrix form:

$$\begin{bmatrix} \sum_{m=1}^M \sum_{n=1}^N \left(\frac{\partial o_1(m, n)}{\partial m} \right)^2 & \sum_{m=1}^M \sum_{n=1}^N \left(\frac{\partial o_1(m, n)}{\partial m} \frac{\partial o_1(m, n)}{\partial n} \right) \\ \sum_{m=1}^M \sum_{n=1}^N \left(\frac{\partial o_1(m, n)}{\partial m} \frac{\partial o_1(m, n)}{\partial n} \right) & \sum_{m=1}^M \sum_{n=1}^N \left(\frac{\partial o_1(m, n)}{\partial n} \right)^2 \end{bmatrix} \begin{bmatrix} h_k \\ v_k \end{bmatrix} = \begin{bmatrix} \sum_{m=1}^M \sum_{n=1}^N (o_k(m, n) - o_1(m, n)) \frac{\partial o_1(m, n)}{\partial m} \\ \sum_{m=1}^M \sum_{n=1}^N (o_k(m, n) - o_1(m, n)) \frac{\partial o_1(m, n)}{\partial n} \end{bmatrix} \quad (9)$$

Equation (9) can be represented in a short form as

$$M \cdot S = V \quad (10)$$

where $S = [h_k \ v_k]^T$,

$$M = \begin{bmatrix} \sum_{m=1}^M \sum_{n=1}^N \left(\frac{\partial o_1(m,n)}{\partial m} \right)^2 & \sum_{m=1}^M \sum_{n=1}^N \left(\frac{\partial o_1(m,n)}{\partial m} \frac{\partial o_1(m,n)}{\partial n} \right) \\ \sum_{m=1}^M \sum_{n=1}^N \left(\frac{\partial o_1(m,n)}{\partial m} \frac{\partial o_1(m,n)}{\partial n} \right) & \sum_{m=1}^M \sum_{n=1}^N \left(\frac{\partial o_1(m,n)}{\partial n} \right)^2 \end{bmatrix},$$

and
$$V = \begin{bmatrix} \sum_{m=1}^M \sum_{n=1}^N (o_k(m,n) - o_l(m,n)) \frac{\partial o_1(m,n)}{\partial m} \\ \sum_{m=1}^M \sum_{n=1}^N (o_k(m,n) - o_l(m,n)) \frac{\partial o_1(m,n)}{\partial n} \end{bmatrix}.$$

Therefore, the estimated registration parameters can be calculated as

$$S = M^{-1}V. \quad (11)$$

To incorporate the case involving larger shift values, we follow the iterative technique of Reference 1. At first, the initial registration parameters are estimated using Eq. (11). Using these estimated shift parameters, $o_k(m,n)$ is shifted so as to closely match $o_l(m,n)$. The resulting image is then registered to $o_l(m,n)$. Using the above mentioned procedure, $o_k(m,n)$ is continuously modified until the registration estimates become sufficiently small. The final registration estimate is obtained by summing all of these partial estimates.

4 Performance Evaluation of the Registration Algorithm

4.1 Effects of Temporal Noise

Temporal noise can be defined as the pixel to pixel variations that are temporally and spatially uncorrelated from frame to frame when the FLIR detector arrays is observing the scene¹⁰⁻¹². Temporal noise changes with each new frame and generally appears as salt and pepper noise. It may include various sources such as readout noise, shot noise, quantization errors, etc. which can be effectively modeled as Gaussian white noise.

To evaluate the effects of spatio-temporal noise, a set of 9 simulated microscanned images or frames was generated from a 450×450 image data collected on a real-time infrared imaging system as shown in Fig. 4. The FLIR camera, manufactured by Mitsubishi, uses a 512×512 Platinum Silicide Schottky Barrier Charge Sweep detector array with 26×20 micrometer center-to-center spacing, 16×12 millimeter chip area, 13.3×10.2 millimeter image area, and 39% fill factor. The sensor has a F# 1.2, 200 millimeter focal length providing 3.5° horizontal and 2.75° vertical field of view with a response in the 3 to 5 micron wavelength band. This detector array has 1.3% quantum efficiency at 3 micrometers, 0.4% quantum efficiency at 4 micrometers and 0.05% quantum efficiency at 5 micrometers. The information about the detector and lens configuration was used in the formation of the system MTF.

This 450×450 image was blurred by the system MTF and sampled by a 150×150 detector array where each detector was simulated with a 3×3 kernel which returns the average of all pixels that fall upon it. The resulting low resolution image is shown in Fig. 5. The set of 9 low resolution images was generated using a prespecified microscan shift pattern. Then Gaussian white noise was added to each frame to form a signal-to-noise ratio (SNR) over the interval of .2 to 20. A low resolution frame with added noise for $\text{SNR} = 10$ is shown in Fig. 6. It should be mentioned that for a given value of SNR, the realization of noise for each frame was different from other frames. The low resolution frames were then input into the shift registration algorithm which calculated the shifts. The mean absolute error (MAE) between the shifts calculated by the registration algorithm and the original programmed shifts was then calculated for each value of the SNR. Fig. 7 shows a plot of the MAE as a function of the SNR. From Fig. 7, it is evident that the MAE of the 9 horizontal and 9 vertical registered shifts was approximately 0.1 of a whole low resolution pixel for $\text{SNR} = 1$. The MAE goes down further for higher values of SNR and always remains below 0.05 when the SNR is increased over 2. From these results, we conclude that the shift registration technique is very robust in estimating the horizontal and vertical registration parameters under extremely low SNR conditions ($\text{SNR} = 1$). This performance may be attributed to the uncorrelated nature of the noise, and the least squares solution provided by the algorithm.

4.2 Effects of Fixed Pattern Noise

Fixed pattern noise, also known as spatial noise, can be defined as the pixel to pixel variations that are temporally correlated from frame to frame when the infrared detector arrays is observing the scene¹⁰⁻¹². Fixed pattern noise arises because the calibration and equalization models can not adequately model actual response, or the drift of the sensors is beyond the capability of practical recalibration schemes. Fixed pattern noise appears as a pattern that remains fixed on the monitor as the scene passes by. In general, fixed pattern noise in scanning arrays will have the appearance of streaks or bands depending on the number of detectors clustering in the process. The pattern of the noise may be fixed, or it may vary between images. There are many sources of fixed pattern noise such as pixel $1/f$ noise, array $1/f$ noise, nonlinear pixel nonuniformities, additive nonuniformities, spectral nonuniformities etc. The pixel nonlinearities are usually corrected at calibration time but residual noise persists away from the calibration points. Pixel $1/f$ noise causes each pixel to drift with respect to the other pixels on the array in a spatially uncorrelated manner. Array $1/f$ noise is related to instability external to all the pixels and usually introduce a pattern onto the image where the pattern itself changes very little in time, but the magnitude of the pattern is variable. Spectral nonlinearities are pixel to pixel variations in the spectral response of the detector and can be corrected by matching the chromatic content of the calibration with that of the image.

Thus the fixed pattern noise is very dependent on the type of sensor and is difficult to model. To obtain insight into the effects of fixed pattern noise, it is assumed that the fixed pattern noise remains constant during the time it takes to acquire the image frames. In addition, the fixed pattern noise is assumed to be Gaussian white noise.

The 450×450 image used in the previous section was blurred by the system MTF and sampled by a 150×150 detector array where each detector was simulated with a 3×3 kernel which returns the average of all pixels that fall upon it. A set of 9 low resolution images was generated using a predefined microscan shift pattern. Then Gaussian white noise was added to each frame to form an SNR over the interval of .2 to 20. A low resolution frame with added noise for SNR = 20 is shown in Fig. 8. It should be mentioned that for a given value of SNR, the realization of noise for

each frame was the same. The low resolution frames were then input into the shift registration algorithm which calculated the shifts. The MAE between the registered shifts and the programmed shifts was then calculated and plotted as a function of the SNR as shown in Fig. 9. From Fig. 9, it is evident that the MAE of the 9 horizontal and 9 vertical registered shifts was approximately 0.1 of a whole low resolution pixel for $\text{SNR} = 5$. The MAE goes down further for increased values of the SNR. From the aforementioned results, we notice that the fixed pattern noise has a more deteriorating effect on the shift registration technique when compared to the effect of temporal noise discussed in the previous subsection.

4.3 Effects of Aliasing

An standard infrared imaging sensor uses a fixed two-dimensional detector array which corresponds to fixed spatial frequency determined by the detector pitch or spacing. According to the Nyquist theorem, aliasing will occur when sampling a scene containing spatial frequencies exceeding half of this sampling frequency. Because most natural scenes are not band limited, aliasing resulting from undersampling can significantly degrade the utility and quality of the image. Accordingly, in this subsection, we investigate the effects of aliasing on the shift registration algorithm.

The 450×450 high resolution image shown in Fig. 4 was blurred by the detector transfer function. Then four different image sequences were constructed using 25%, 36%, 56.25% and 100% fill factors. Each of these sequences consists of 9 low resolution frames generated using a preselected microscan shift pattern. Then Gaussian white noise was added to each frame to form an SNR over the interval of .2 to 20. It should be mentioned that for a given value of SNR, the realization of noise for each frame was the same i.e., fixed pattern noise was used. The low resolution frames were then input into the shift registration algorithm which calculated the shifts. The MAE between the registered shifts and the programmed shifts was then calculated and plotted as a function of the SNR as shown in Fig. 10. From Fig. 10, it is evident that the MAE of the 9 horizontal and 9 vertical registered shifts was approximately 0.1 of a whole low resolution pixel for $\text{SNR} = 5$ if the fill factor is 100%. On the other hand, the MAE of the registered shifts was approximately 0.22 of a whole low resolution pixel for $\text{SNR} = 5$ if the fill factor is 25% - an

increase of error by a factor of 2.2. The same trend of results were obtained for the other fill factors. Therefore, as the fill factor is decreased the MAE increases rapidly which implies that the MAE of the shifts registered by the algorithm increases with the increase in aliasing. Furthermore, with the increase in SNR the rate of decrease in the MAE with smaller fill factors (such as 25%) is much lower than that with larger fill factors (such as 100%). Thus, as the aliasing increases, the shift registration technique becomes more prone to increased MAE.

4.4 Effects of Image Size

As mentioned earlier, the shift registration algorithm is very robust in estimating the horizontal and vertical (i.e., x-y) registration parameters under extremely low SNR conditions. This is mainly due to the uncorrelated nature of the noise, and the least squares solution provided by the registration algorithm. Notice that the least squares solution effectively averages over all the pixels in the images being compared and consequently the noise effects tend to average out. In cases where the image size is small, it is expected that the least squares solution will be less accurate and the registration algorithm may not yield very accurate results.

To investigate the effect of image size on the shift registration algorithm, we considered the 450×450 image used earlier which was blurred by the system MTF and sampled by a 150×150 detector array where each detector was simulated with a 3×3 kernel which returns the average of all pixels that fall upon it. A set of 9 low resolution images was generated using a prescribed microscan shift pattern where each frame consists of 150×150 pixels. Then, 13 additional sequences of frames were generated by gradually reducing the number of pixels from each frame (such as 140×140 , 130×130 , 120×120 , ..., 30×30 , and 20×20). Each set of low resolution frames were then input into the shift registration algorithm which calculated the shifts for each case. The MAE between the shifts calculated by the registration algorithm and the original programmed shifts was then calculated and the peak MAE for each set of frames was recorded. Fig. 11 shows a plot of the MAE as a function of the frame size where the numbers along the X-axis indicate the frame size (e.g., a number 20 indicates the size of each frame is 20×20). From Fig. 11, it is evident that the MAE of the 9 horizontal and 9 vertical registered shifts was less than 0.1 of a

whole low resolution pixel for sequences with frame sizes 30×30 and above, and the MAE was less than 0.04 of a whole low resolution pixel for sequences with frame size larger than 40×40 . It may be mentioned that the aforementioned results were obtained in the absence of noise.

Next, we investigated the effect of image size on the shift registration algorithm in the presence of fixed pattern noise. We considered five different sets of frame sequences with frame size equaling 100×100 , 80×80 , 60×60 , 40×40 , and 20×20 respectively. Then Gaussian white noise was added to each frame to form an SNR over the interval of .2 to 20. It should be mentioned that for a given value of SNR, the realization of noise for each frame was the same. The low resolution frames from each set were then input into the shift registration algorithm which calculated the shifts. The MAE between the registered shifts and the programmed shifts was then calculated and plotted as a function of the SNR as shown in Fig. 12. From Fig. 12, it is evident that as the number of pixels in each frame of a sequence decreases, the MAE increases and the rate of increase is higher with smaller number of pixels per frame.

4.5 Effects of Image Contrast

To investigate the effect of image contrast on the registration algorithm, various contrast imagery produced from simulations of the identical scene at different times of day and year are needed. Figure 13 shows a real 450×450 infrared image obtained at the Wright laboratory using the $3\text{-}5\text{ }\mu\text{m}$ thermal imaging system discussed earlier. The histogram of this image is shown in Fig. 14a. Since it is important to adjust the contrast of the imagery to represent the contrast in typical infrared imagery as closely as possible, five additional images were created from this image by gradually modifying the image histogram as shown in Figs. 14b through 14f. A set of 9 low resolution images was generated from each of these images (corresponding to Figs 14a through 14f) using *a priori* selected microscan shift pattern. Then Gaussian white noise was added to each frame to form an SNR over the interval of .2 to 20. It should be mentioned that for a given value of SNR, the realization of noise for each frame was the same. The low resolution frames from each set were then input into the shift registration algorithm which calculated the shifts. The MAE between the registered shifts and the programmed shifts was then calculated and plotted as

a function of the SNR with image contrast and variance as parameters as shown in Fig. 15. We have defined image contrast as

$$Contrast = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}. \quad (12)$$

where I_{max} refers to the maximum value of image intensity while I_{min} refers to the minimum value of image intensity. From Fig. 15, it is obvious that for a low contrast image (i.e. a typical infrared image), the MAE is considerably high. However, the variation in image contrast (corresponding to the image histograms of Figs. 14a through 14f) has little impact on the registration algorithm. The MAE was found to decrease gradually with the increase of SNR.

5 Conclusion

The image registration algorithm proposed by Irani and Peleg¹ has been found to be an extremely fast technique when compared to the alternate algorithms. This algorithm has been demonstrated to work with non uniformity-corrected, aliased imagery collected from a laboratory 3- to 5- μ m thermal imaging system. However, the performance of this algorithm over all imaging conditions is unknown. Accordingly, in this research project, we have evaluated the performance of this registration algorithm under various levels of degrading factors including the fixed-pattern spatial noise, temporal noise, aliasing, contrast, and image size. Plots of registration error versus degradation measure has been provided in order to indicate the general trend of effects of each of these degrading factors. Among these degrading factors, the temporal noise has been found to have the least deteriorating impact on the algorithm while the other factors introduce varying levels of errors in estimating the translational shifts. It was also observed that the image registration error or the MAE decrease gradually with the increase in fill factor (i.e., a decrease in aliasing). Furthermore, the image registration error also decreases with the increase in image size.

Acknowledgment

The authors would like to thank the following individuals for their assistance and support: Ken Barnard, Brian Yasuda, John Bognar, Ernie Armstrong, Ed Watson, and Donald Tomlinson.

References

- [1] M. Irani and S. Peleg, "Improving Resolution by Image Registration," *CVGIP: Graphical Models and Image Processing*, Vol. 53, pp. 231-239 (1991).
- [2] E. A. Watson, R. A. Muse and F.P. Blommel, "Aliasing and blurring in microscanned imagery," *Proc. of the SPIE*, Vol. 1689, 242-250 (1992).
- [3] J. C. Gillette, T. M. Stadmler, and R. C. Hardie, "Reduction of Aliasing in Staring Infrared Imagers Utilizing Subpixel Techniques," *Optical Engineering*, Vol. 34, pp. 3130-3137 (1995).
- [4] S. Cain and R. C. Hardie, "Restoration of Aliased Video Sequences via a Maximum-likelihood Approach," *Proceedings of the IRIS Symposium on Passive Sensors*, Monterey, CA, March (1996).
- [5] R. C. Hardie and S. Cain, "MAP Estimation of a High Resolution Image From a Sequence of Undersampled Translationally Shifted Images," Under review, *IEEE Transactions on Image Processing*, February (1996).
- [6] M. S. Alam, J. G. Bognar, B. J. Yasuda and R. C. Hardie, "Infrared Image Registration and High Resolution Reconstruction using Multiple Translationally Shifted Aliased Video Sequences," under review, *IEEE Transactions on Instrumentation and Measurement*, March 1997.

- [7] A. Schaum and M. McHugh, "Analytic Methods of Image Registration: displacement Estimation and Resampling," Naval Research Report 9298, Feb. 28, (1992).
- [8] E. Kaltenbacher and R. C. Hardie, "Infrared Image Registration and High Resolution Reconstruction," *Proceedings of the National Aeronautics and Electronics Conference*, pp. 702-709 (1996).
- [9] R. C. Gonzalez and R.E . Woods, Digital Image Processing, Addison-Wesley, New York (1992).
- [10] J. M. Mooney, "Effect of Spatial Noise on the Minimum Resolvable Temperature of a Staring Sensor," *Applied Optics*, Vol. 30, pp. 3324-3332 (1990).
- [11] K. J. Barnard and E. A. Watson, "Effects of Image Noise on Submicron Interpolation," *Optical Engineering*, Vol. 34, pp. 3165-3173 (1995).
- [12] M. D. Nelson, J. F. Johnson and T. S. Lomheim, "General Noise Processes in Focal Plane Arrays," *Optical Engineering*, Vol. 30, pp. 1682-1700 (1991).



Figure 4. A 450×450 pixel high resolution image collected on a real-time infrared imaging system.

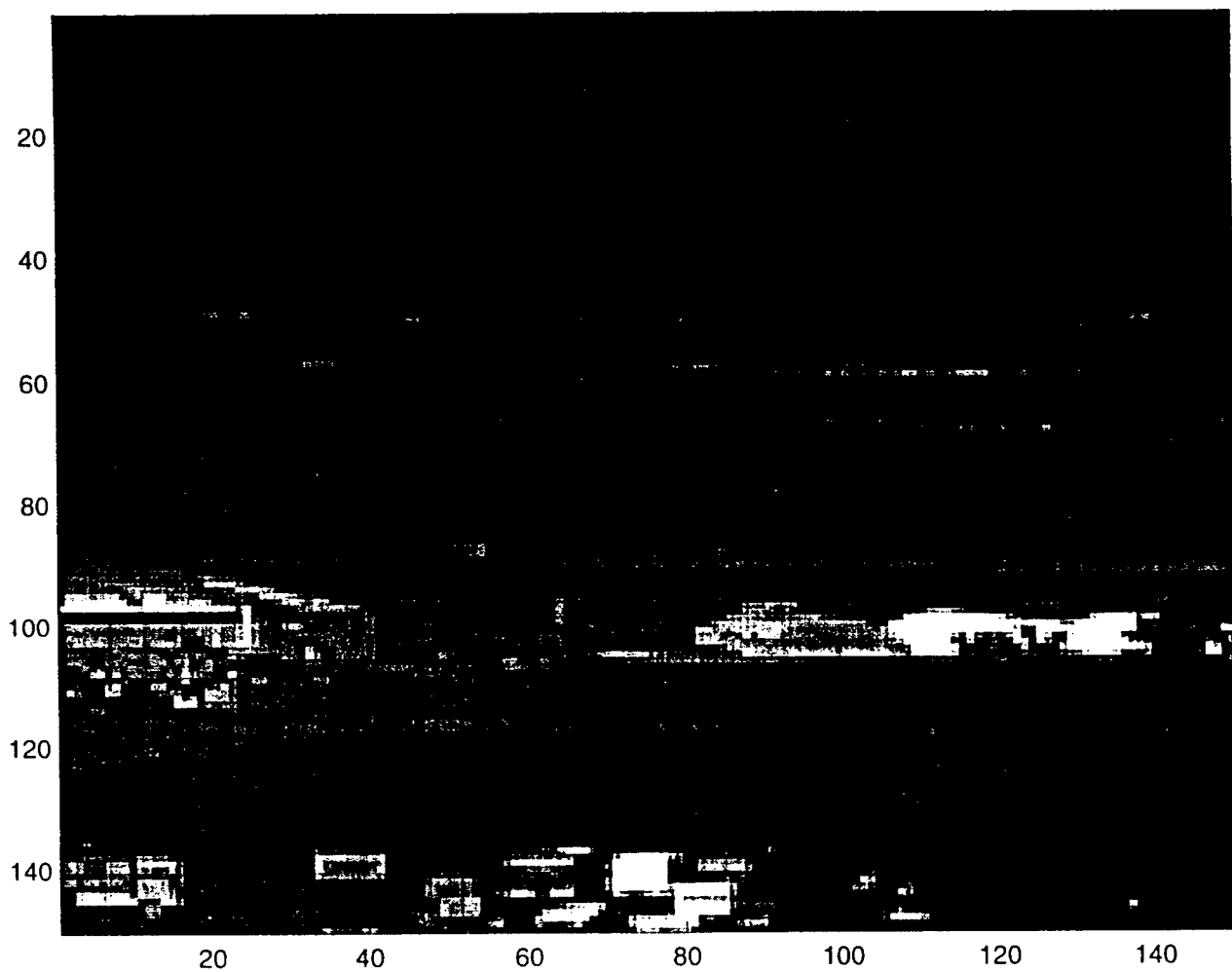


Figure 5. A 150×150 low resolution image collected by the detector array from the 450×450 pixel high resolution image where each detector was simulated with a 3×3 kernel which returns the average of all pixels that fall upon it.



Figure 6. A low resolution frame with added noise for a signal-to-noise ratio (SNR) of 10.

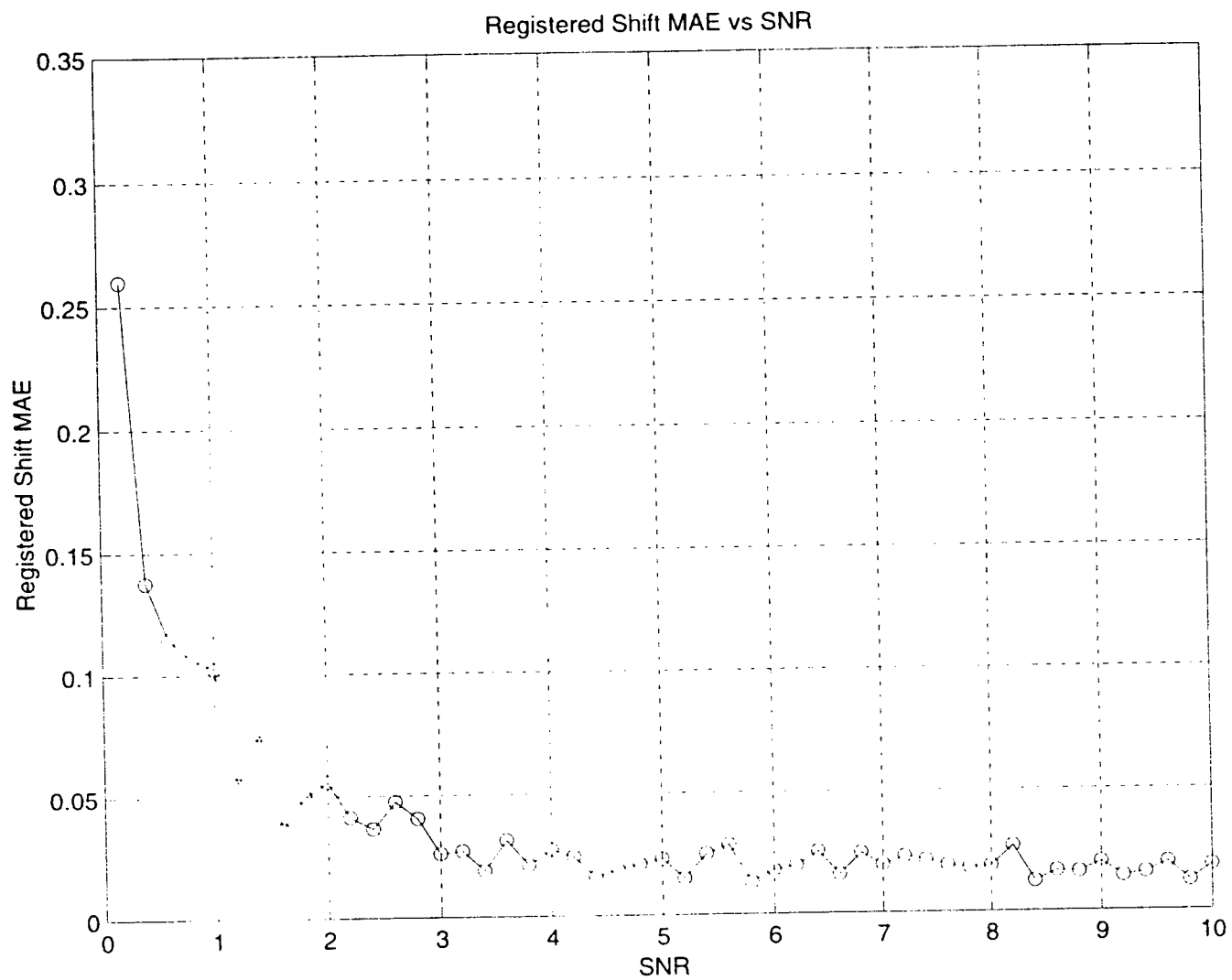


Figure 7. Plot of the mean absolute error (MAE) as a function of SNR showing the effect of spatio-temporal noise on the registration algorithm.

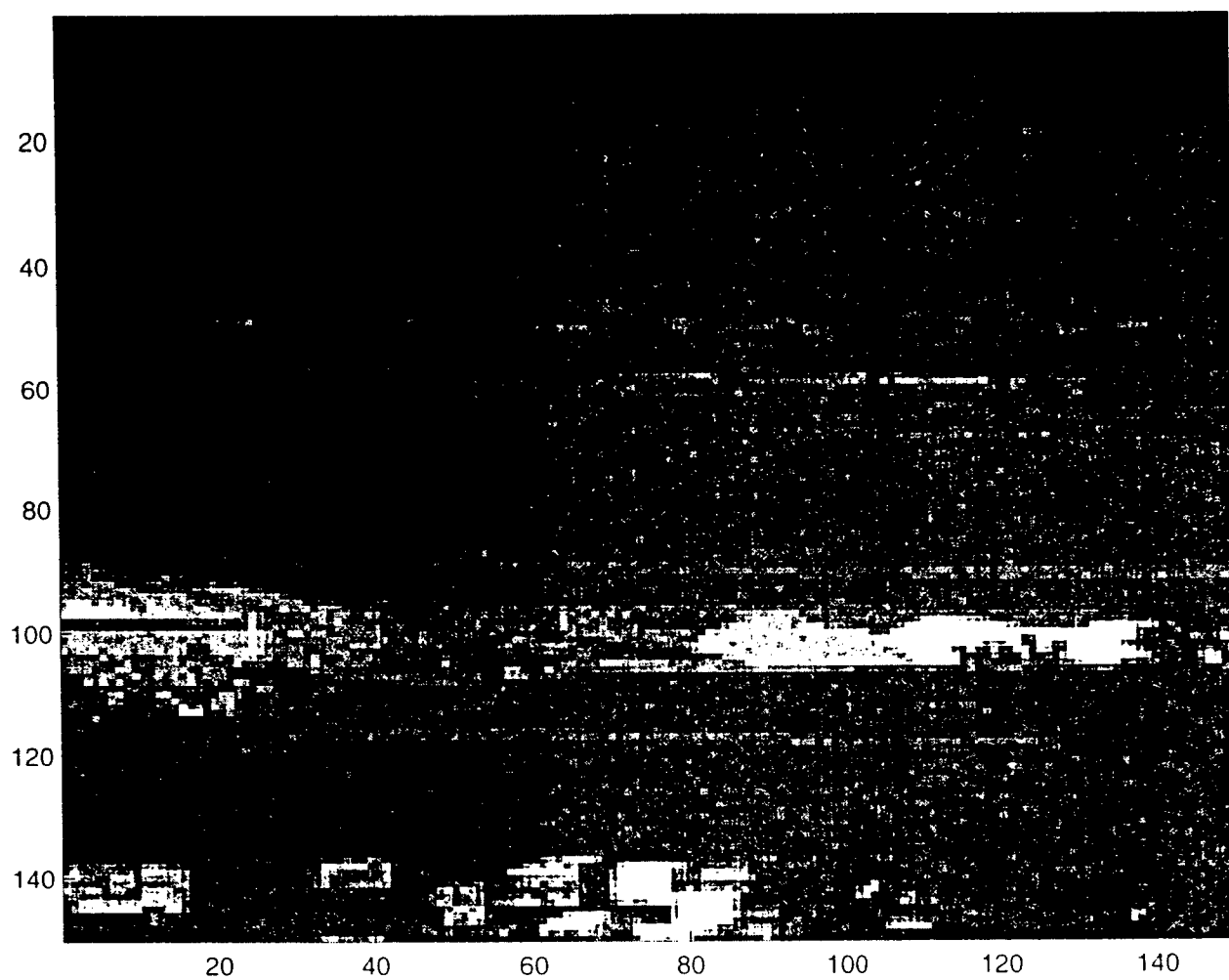


Figure 8. A low resolution frame with added noise for a signal-to-noise ratio (SNR) of 20.

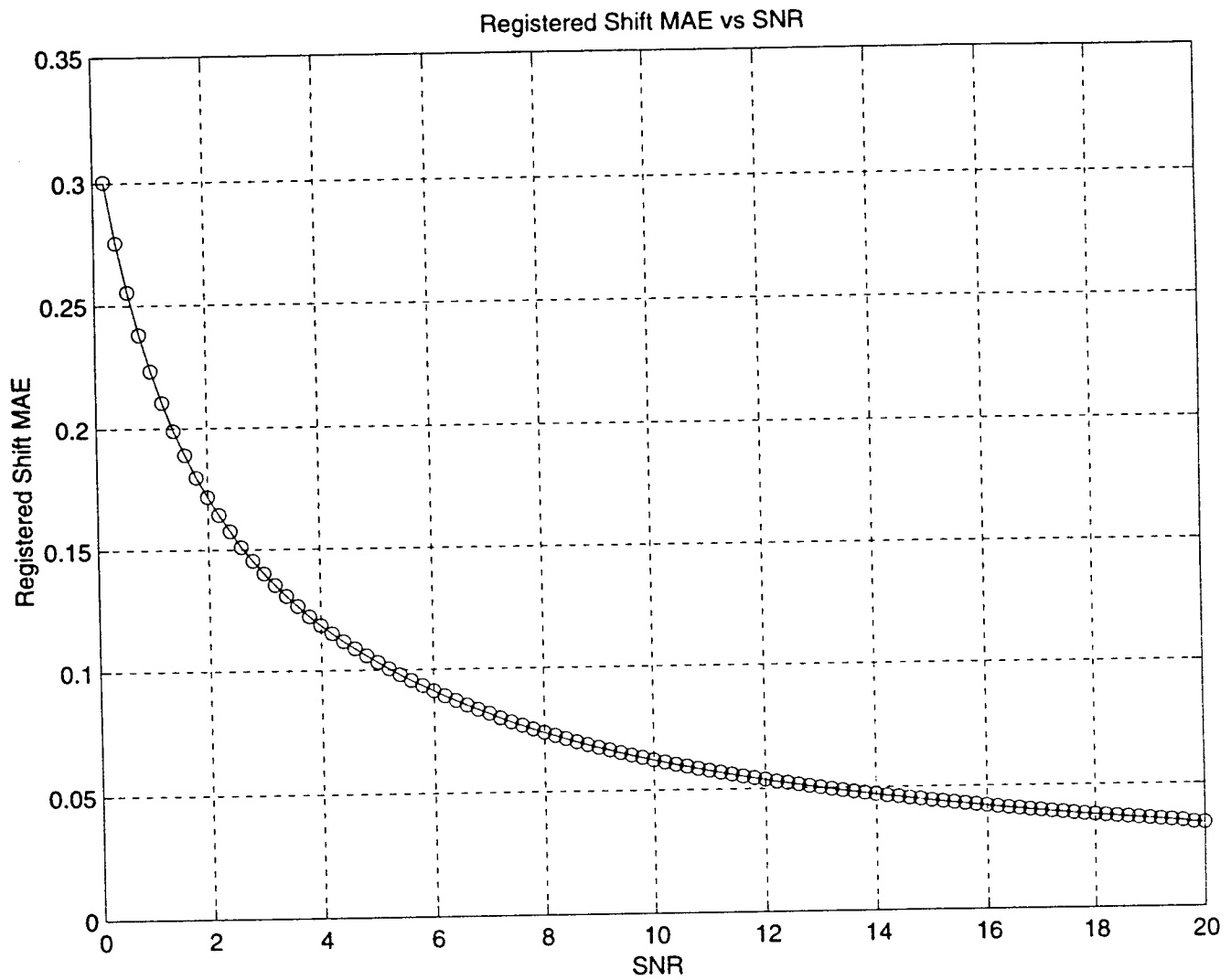


Figure 9. Plot of MAE as a function of SNR showing the effect of fixed pattern noise on the registration algorithm.

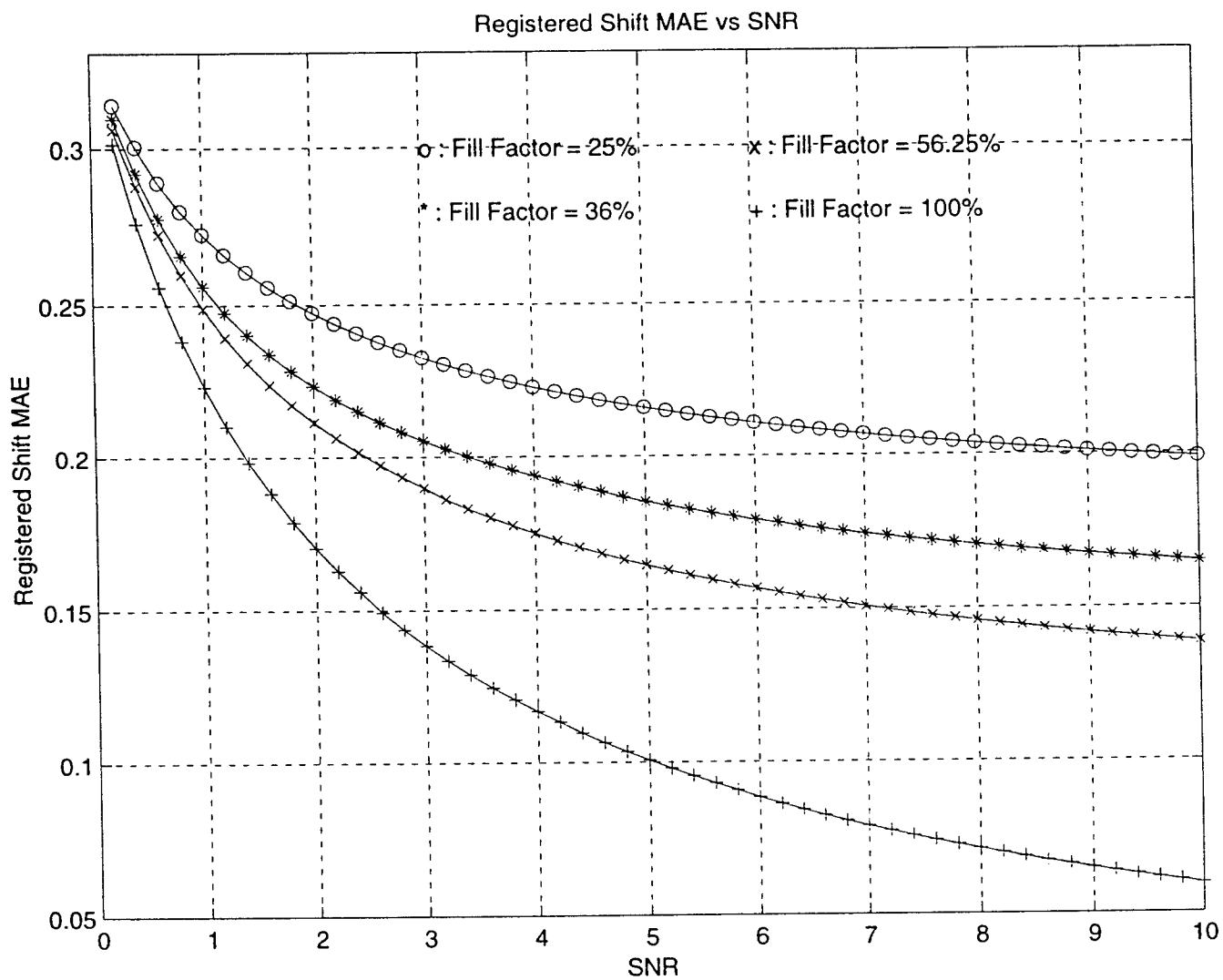


Figure 10. Plot of MAE as a function of SNR with fill factor as a parameter showing the effect of various levels of aliasing on the registration algorithm.

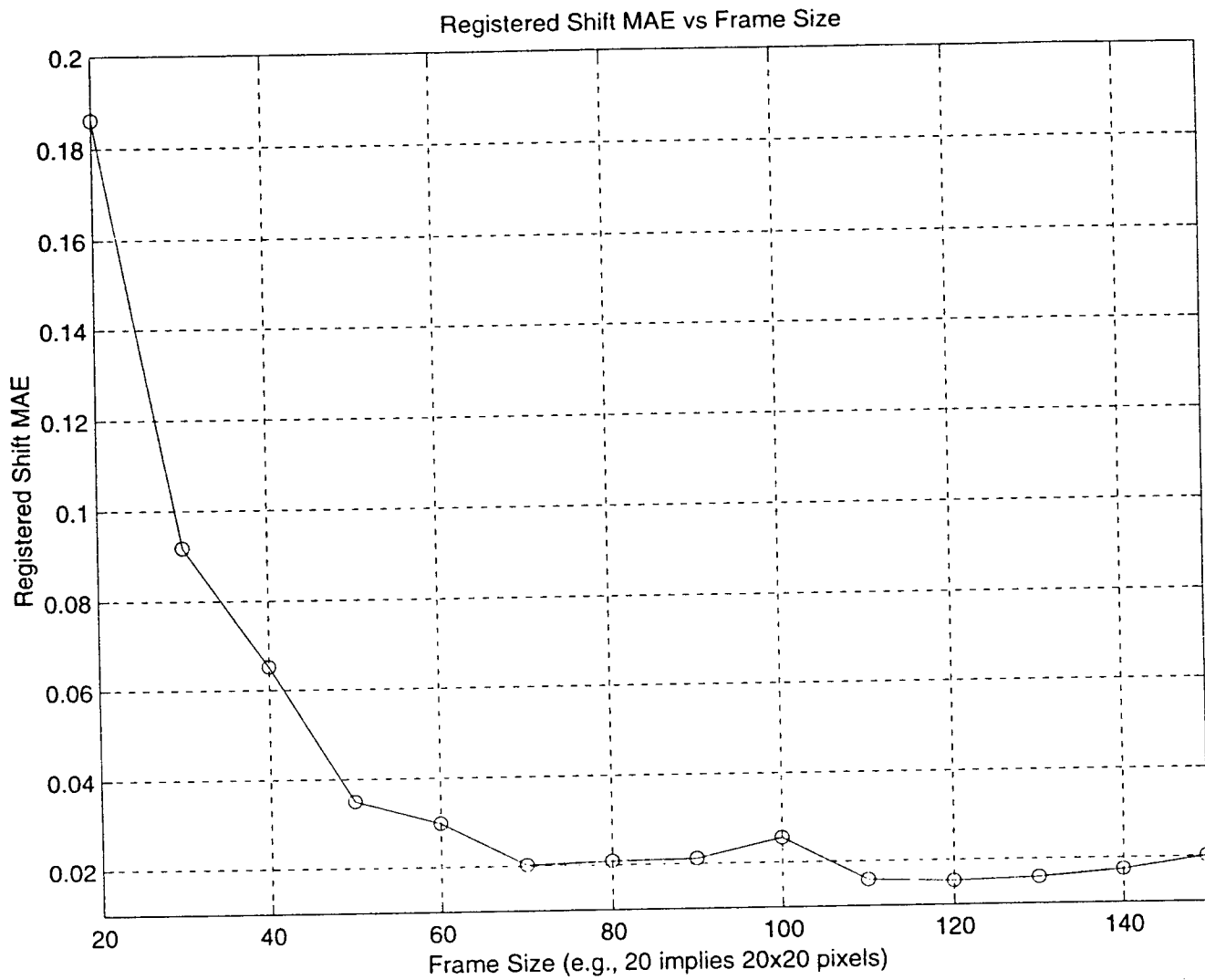


Figure 11. Plot of MAE as a function of low resolution frame size in the absence of noise showing the effect of different frame size on the registration algorithm.

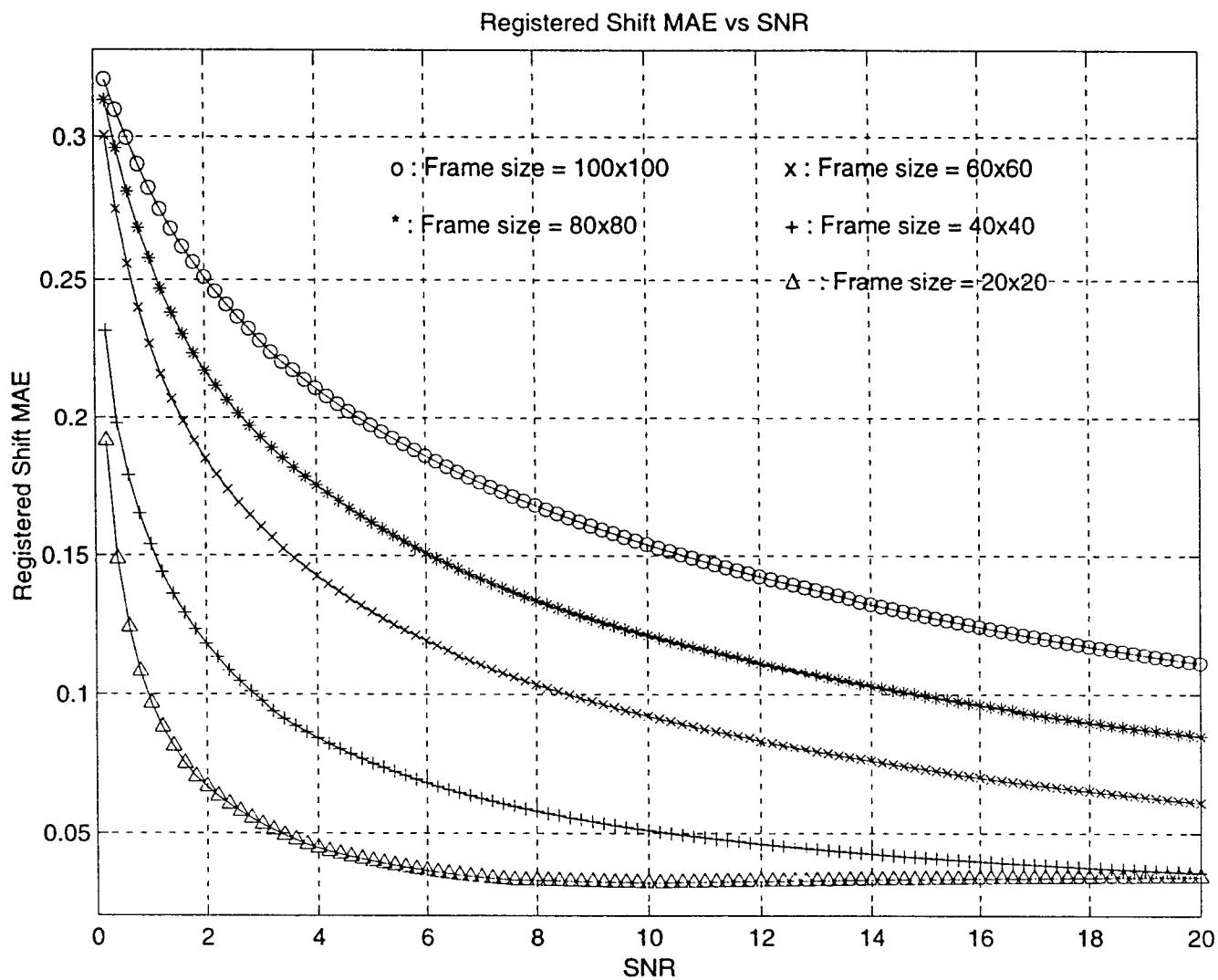


Figure 12. Plot of MAE as a function of SNR with frame size as a parameter showing the effect of different frame size on the registration algorithm in the presence of fixed pattern noise.

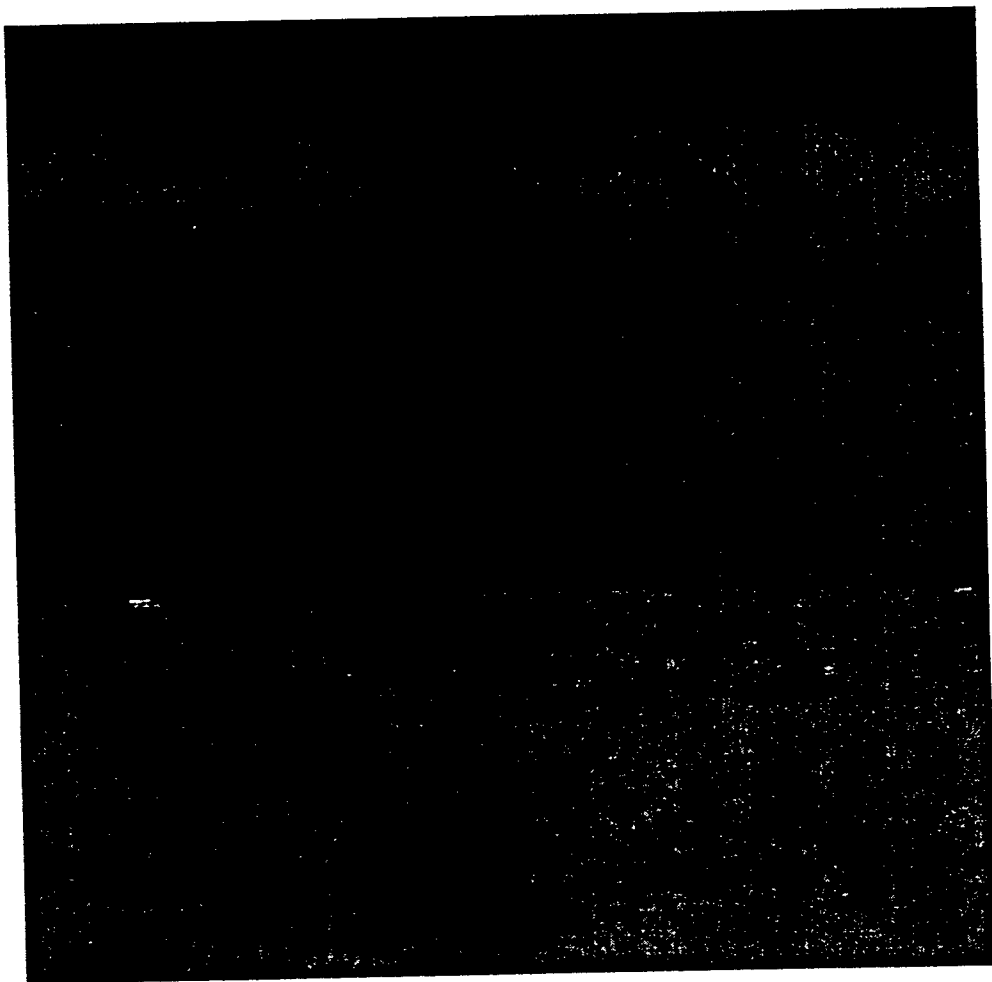


Figure 13. A real 450×450 high resolution infrared image obtained at the Wright laboratory using a 3-5 μm thermal imaging system.

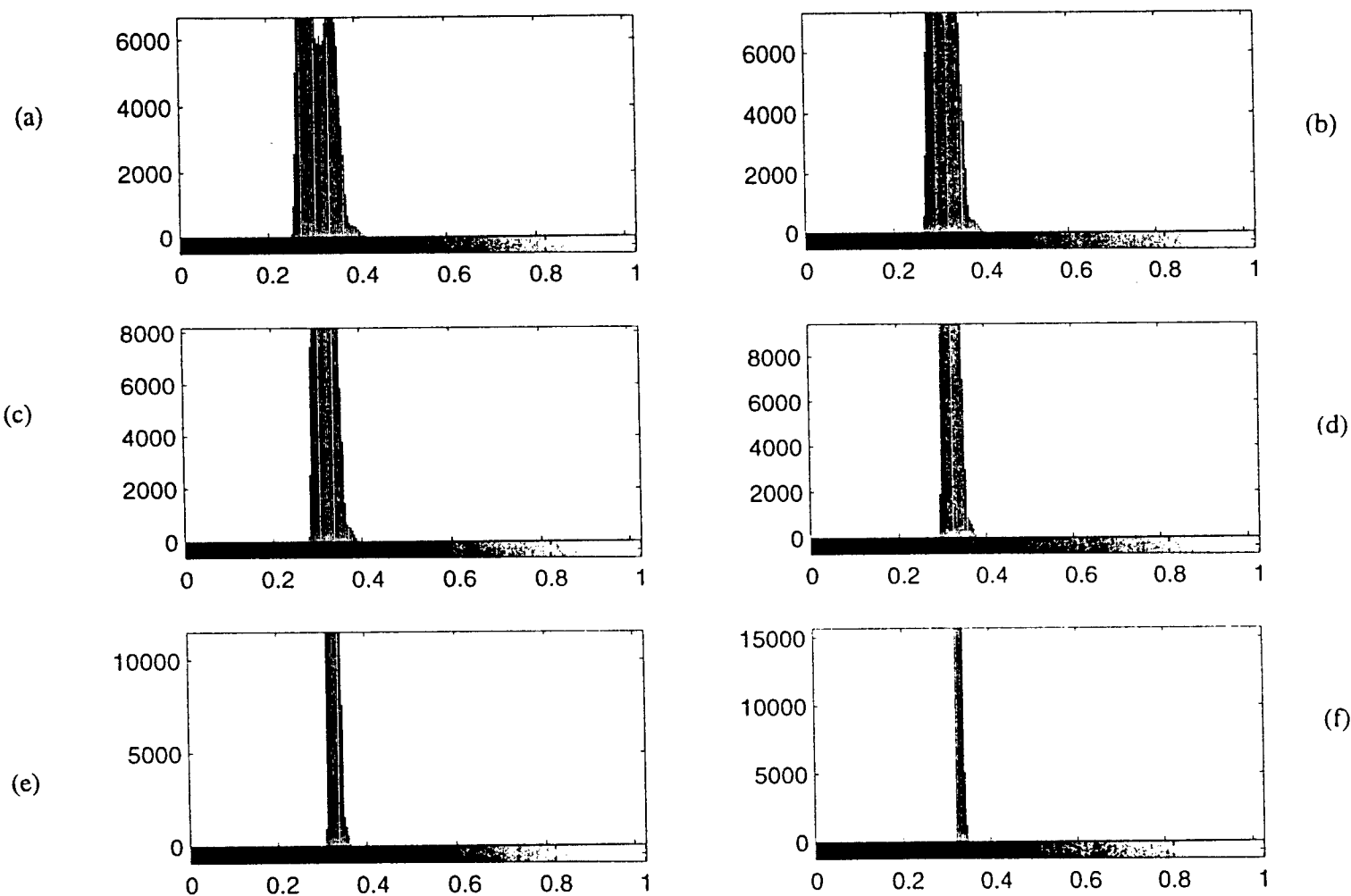


Figure 14. (a) Histogram of the image is shown in Fig. 13, (b)-(f) Histograms of the five additional high resolution images obtained from the image of Fig. 13 after applying various histogram modifications.

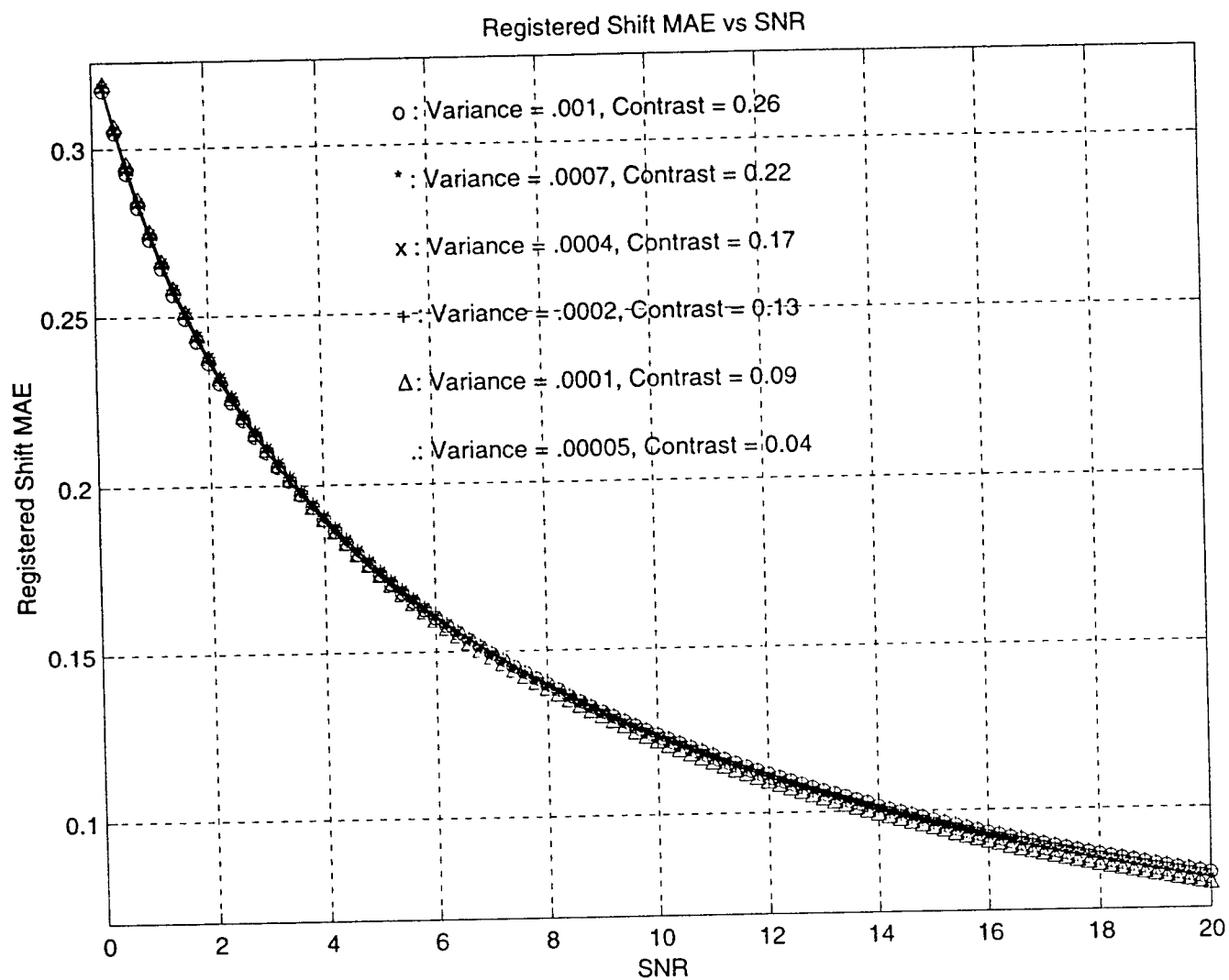


Figure 15. Plot of MAE as a function of SNR image contrast and variance as parameters showing the effect of variation in image contrast on the registration algorithm.

TEXTURE ,MICROSTRUCTURE AND MICROTTEXTURE
OF HOT ROLLED Ti-6Al-4V

Pnina Ari-Gur
Associate Professor
Department of Construction Engineering,
Materials Engineering & Industrial Design

Western Michigan University
Kalamazoo, MI 49008-5064

Final Report for:
Summer Research Extension Program
97-0819
Wright Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Wright Laboratory

March 25, 1998

TEXTURE ,MICROSTRUCTURE AND MICROTEXTURE OF HOT ROLLED Ti-6Al-4V

Pnina Ari-Gur
Associate Professor
Department of Construction Engineering,
Materials Engineering & Industrial Design
Western Michigan University
Kalamazoo, MI 49008-5064

Abstract

The development of microstructure, macrotexture, and microtexture during subtransus hot working of Ti-6Al-4V with two different types of transformed beta starting microstructures (lamellar colony, acicular/martensitic alpha) was studied.

Globularization of the transformed microstructures required heavy rolling reductions or moderate reductions coupled with near transus post-rolling heat treatment. Noticeable macrotexture changes were noted after low reductions, an effect ascribed to the rotations associated with kinking and bending of the lamellar/acicular plates.

Noticeable microtextures, noted in samples with an initial lamellar colony microstructure, persisted through hot rolling suggesting that dynamic globularization does not involve recrystallization. In contrast, hot rolled material with a starting acicular alpha microstructure exhibited weak microtextures and strong macrotextures. The absence of microtexture in these latter cases was explained on the basis of multiple transformation variants within each colony/prior beta grain following the beta-annealing-and-water-quenching process used to obtain the microstructure.

TEXTURE ,MICROSTRUCTURE AND MICROTTEXTURE OF HOT ROLLED Ti-6Al-4V

Pnina Ari-Gur

1. Introduction:

Ti-6Al-4V is a high-strength light-weight material that has been in extensive use for aerospace applications. Its processing is difficult -- it is not workable at low temperatures, while at temperatures above 75% of the solidus temperature, its developing microstructure is hard to control [1].

At room temperature and up to 990°C, Ti-6Al-4V has a dual phase ($\alpha+\beta$) structure. The α -phase (HCP), the dominant phase throughout this temperature range, is highly anisotropic in nature. The second phase (β) is BCC, and is the single phase present above the transus temperature (990°C). A number of investigators have examined the kinetics of the globularization of alpha/beta titanium alloys during subtransus hot working [2-7]. Most of these efforts have focused on monotonic deformation involving modes of loading such as uniaxial compression, tension, or simple torsion. However, most mill processing relies on multipass operations involving small reductions per pass and intermediate reheating and/or annealing steps. Thus, it may be difficult to correlate laboratory and production observations.

Although globularized alpha microstructures may be free of mechanical texture, it is well known that a considerable degree of crystallographic texture may be produced [8,9]. Because of the anisotropic nature of the hexagonal alpha phase, such textures may affect both first- and second-tier service properties such as elastic modulus and

fatigue resistance [8,10]. It is also becoming apparent now that the development of microtexture, or large regions of similarly oriented alpha grains, may impact second-tier properties such as dwell fatigue resistance [11].

The aligned α colonies can be traced back to the original β (BCC) grain that existed above the transus temperature. "Breaking" the original microstructure, as well as elimination of the micro-texture (local aligned colonies), are what thermo-mechanical processing [1] is aimed at.

The goal of this work was to study the effects of initial microstructure, hot-rolling parameters (temperature and percent reduction), and post-rolling heat treatments on the evolution of microstructure, crystallographic texture, and micro-texture.

2. Experimental Procedures

2.1 Material

The material used in this investigation was received as 30 mm thick hot-rolled plate of Ti-6Al-4V. Its measured composition in weight percent was 6.09 aluminum, 3.95 vanadium, 0.18 iron, 0.19 oxygen, 0.02 carbon, 0.01 nitrogen, and 0.010 hydrogen, with the balance being titanium. The beta transus temperature was 995°C.

Samples measuring 76 x 102 mm were cut from the as-received plate and given one of two heat treatments - beta annealed plus controlled cooled or beta annealed plus water quenched. The former treatment hereafter referred to as BACC comprised 940°C/20 min. + 1065°/15 min. + 940°C/5 min. + 815°C/15 min. + air cool. The

microstructure developed by this means consisted of a mixture of colony and basketweave alpha with the prior beta grain boundaries decorated with a layer of alpha (Figure 1a). The prior-beta grain size was approximately 500 μm , and the colonies were approximately 250 μm in size. Samples given the beta-annealed-plus-water-quenched treatment (subsequently referred to as BAWQ) were heated at 940°C/20 min. + 1065°C/15 min. and then water quenched. This treatment produced a fine acicular (martensitic alpha) microstructure with a similar prior-beta grain size but no grain-boundary alpha (Figure 1b).

2.2 Thermomechanical Processing (TMP)

The BACC and BAWQ samples were processed using various rolling or rolling-plus-heat-treatment sequences to assess the effect of thermomechanical treatments on the breakdown of the transformed-beta type microstructures and the development of texture and microtexture. Rolling was done subtransus at temperatures of 815, 925, or 970°C. Samples were furnace heated for 40 minutes prior to rolling and reheated 3 minutes between each pass. The rolling speed was 25 cm/s, and the reduction per pass was 12 percent (two higher rolling temperatures) or 5 percent (lowest rolling temperature). Rolling was continued until a total thickness reduction of either 44 or 77 percent (corresponding to thickness strains of 0.57 and 1.47 or effective strains of 0.66 and 1.69, respectively) had been achieved. Following rolling, all samples were air cooled to simulate production practice.

To determine the effect of static versus dynamic processes on globularization and or recrystallization of the transformed beta microstructures and the concomitant texture and microtexture evolution, samples cut from the hot rolled sheets were given either a

mill anneal or a higher temperature solution annealing treatment. The mill anneal was conducted at 700°C for one hour followed by air cooling; this treatment did not lead to noticeable changes in microstructure but was necessary to relieve residual stresses, etc., for the successful application of the orientation imaging technique for the determination of microtexture. Solution heat treatments were conducted for two hours at temperatures between 815°C and 970°C followed by air cooling. Details of the post-processing heat treatments are given in Table 1.

Table 1: Solution Annealing Treatments

Group Number	Temperature	Time	Cooling
1	Same as rolling	2 hours	Air
2	870°C	2 hours	Air
3	955°C	2 hours	Air

2.3 Post-TMP Characterization

Following rolling or rolling plus heat treatment, microstructure, macrotexture, and microtexture were determined. Microstructure was quantified through optical metallography on thickness cross-sections cut parallel to the rolling direction. Crystallographic texture and microtexture were determined for the alpha phase. For the determination of global texture (or so-called macrotexture), standard (0002) pole figures were measured via a back reflection technique using copper K_{α} radiation. Samples for such measurements were prepared by grinding and polishing 25 x 25 mm coupons to the midplane. For ease of visualization of the primary macrotexture components, the pole figures were processed through the program popLA [12] to obtain graphic representations of the components. The local (spatial) variation of texture (or microtexture) was obtained by analysis of electron backscatter patterns (EBSP) in a

scanning electron microscope [13,14] For this purpose, thickness cross-sections parallel to the transverse direction were cut and prepared metallographically. A region close to the rolled surface and one at the midplane were analyzed. In each case, the area sampled was approximately 4 mm^2 over which 40,000 discrete measurements were made. These microtexture data were represented in the form of icosahedral and total Eulerian plots delineating the orientation of the (0002) basal poles so determined. Icosahedral plots are generated by dividing the spherical projection of the offset angle between the (0002) poles and the fixed sample axes into 10 regions. Each region is represented by a different color, so identical color indicates similar orientations. Points where indexing is not possible ("no hit") are white. Total Eulerian plots are essentially similar to the icosahedral plots, except that there are thousands of color possibilities instead of 10. The background for the total Eulerian plots is black.

3. Results and Discussion

3.1 Microstructure

The as-rolled microstructures of the BACC and BAWQ specimens are shown in Figures 2 and 3. For the BACC material, rolling to the lower reduction $\bar{\varepsilon} = 0.66$ gave rise to relatively little globularization. The majority of the microstructure was characterized by bent, higher aspect-ratio alpha plates (Figure 2b). In contrast, rolling to a higher reduction $\bar{\varepsilon} = 1.69$ yielded a sizable proportion of alpha grains with aspect ratios of less than 3:1 at all three rolling temperatures (815, 925, 970°C) as shown in Figures 2 a, c, e. By and large, the effect of reduction and rolling temperature on the globularization of BAWQ specimens (Figure 3) was very similar to the observations for the BACC samples. This is particularly evident for samples rolled at 970°C (Figures 3 c, d vs. Figures 2 d, e); the resemblance between the two sets of microstructures at 970°C may be ascribed to continuous coarsening of the acicular alpha plates which

occurs very rapidly during preheating at this temperature [15]. The coarsening kinetics are somewhat slower at 925°C. Thus, the globularized microstructure in BAWQ samples rolled at 925°C (Figure 3b) was somewhat finer than that in the corresponding BACC condition (Figure 2c).

The effect of post-rolling heat treatment on the degree of globularization is illustrated in Figures 4, 5 and 6. Little increase in globularization is evident after two-hour heat treatments at 870°C (Figures 4 and 5) when compared to the as-rolled microstructure (Figure 2). On the other hand, samples which were only lightly rolled at 925°C were globularized to a large extent after a two-hour heat treatment at 925°C (Figures 6b, c).

3.2 *Macrotexture Development*

The (0002) pole figures measured for the as-heat-treated BACC and BAWQ samples are shown in Figure 7. Both materials exhibited a pronounced macrotexture. In times random units, the BACC had three principal texture components - TD (6X), a so-called 45° component (8X), and RD (2X) - and the BAWQ had two components - TD (16X) and 45° (10X), where RD and TD refer to rolling direction and transverse direction, respectively. Because of the large beta grain size and hence moderate number of colonies inspected to obtain the pole figures, the textures may seem stronger than they actually are. Therefore, the relative intensities may be somewhat approximate and should thus be interpreted in semi-quantitative terms. Nevertheless, there is no doubt that annealing the Ti-6Al-4V plate in the beta field did not result in a random macrotexture.

The present beta-annealed pole figures (Figure 7) can be compared to those measured by previous investigators. Moustahfid, et al. [16] utilized a Ti-6Al-4V sheet material which had been processed through a series of thermomechanical treatments including a final cold rolling step and mill anneal. Prior to beta annealing, this material exhibited a strong basal texture ((0002) poles parallel to the sheet normal). After beta annealing, the alpha phase exhibited two very strong components in the (0002) pole figure, an RD component and a component between the sheet normal and RD. Moustahfid, et al. [16] concluded that some complex form of texture variant selection must have occurred, because the six possible variants characterizing the alpha-to-beta transformation and the twelve variants for the beta-to-alpha transformation would not have produced such strong textures in the absence of such a variant selection. In contrast to the findings of Moustahfid, et al., Peters and Luetjering [17] found that beta annealing followed by water quenching eliminated the alpha-phase texture of Ti-6Al-4V plate material that had been produced by beta forging.

The alpha-phase macrotextures of the BACC and BAWQ samples after hot rolling are summarized in Figure 8. These (0002) pole figures reveal that the major texture components were the TD and RD ones, whereas basal and 45° textures were found in only a few cases. The trends are quantified by the data in Table 2, which shows the predominance of the TD component in all cases. The TD component in the rolled materials may be thought to have evolved from the TD component present in the as-heat treated materials. It is also possible that the TD component could come from the decomposition of hot worked beta phase as noted by Gey, et al. [18].

The results in Table 2 also reveal the noticeable absence of a 45° texture component (except for low reduction hot rolling of BAWQ material at 925°C). As

noted above, both the BACC and BAWQ samples had strong 45° components in the as-heat-treated condition. The absence of the 45° component is even more remarkable in view of its elimination after relatively low deformation levels. This observation may be explained on the basis of the kinking and bending of lamellar plates at relatively low strains noted in Section 3.1. Such kinking could give rise to noticeable plate rotations and thus texture changes.

Table 2. Pole Figure Measurements of Texture Components Developed in Ti-6Al-4V

Material	Condition	Rolling Temperature (°C)	Reduction*	Texture Components
BACC	As Heat Treated	----	----	TD (6X), 45° (8X), RD (2X)
BACC	Hot Rolled	815	High	TD (9X), Basal (2X)
BACC	Hot Rolled	925	Low	TD (6X), RD (4X)
BACC	Hot Rolled	925	High	TD (8X), RD (3X)
BACC	Hot Rolled	970	Low	TD (8X), RD (6X)
BACC	Hot Rolled	970	High	TD (8X), RD (3X)
BAWQ	As Heat Treated	----	----	TD (16X), 45° (10X)
BAWQ	Hot Rolled	925	Low	TD (8X), 45° (2X)
BAWQ	Hot Rolled	925	High	TD (8X), RD (3X)
BAWQ	Hot Rolled	970	Low	TD (4X), RD (4X)
BAWQ	Hot Rolled	970	High	TD (9X), RD (4X)

* Low reduction = 44 pct, high reduction = 77 pct.

The development of RD texture components in the as-hot rolled materials (Table 2) may also be controlled by texture changes associated with lamellar plate kinking and bending. For example, the BACC samples which had a 2X RD component in the as-

heat treated condition, showed a noticeably stronger RD component after light reductions. At the higher deformation levels, however, the transient nature of RD texture was manifested by a decreasing sharpness while the TD component remained relatively sharp.

The almost total absence of the basal texture components in the rolled Ti-6Al-4V samples contrasts sharply with the observations of Peters and Luetjering [17], who measured noticeable components of this type following subtransus hot rolling over a wide temperature range. This difference may be at least partly due to different textures prior to hot rolling. Further work is needed to establish the exact source of the difference in macrotexture.

The effect of post-rolling heat treatment on the macro-texture is illustrated in Figures 9 and 10. The TD component is still evident after the heat treatment and in addition the 45° component reappears.

3.3 Microtexture Evolution

The EBSD measurements to quantify the alpha phase microtexture provide additional insight into the interdependence of microstructural and textural evolution. The microtexture results to be discussed pertain solely to the midplane region of the plate/sheet samples, i.e., the location at which the pole figures had been measured.

Figure 11 summarizes the alpha-phase microtextures of the as-heat treated BACC and BAWQ materials in both icosahedral and total Eulerian forms. The BACC material (Figures 11a, b) showed very distinct patches of various colors. These patches are

approximately 250 μm in size and can be correlated to the colony features in the optical microstructure (Figure 1a). The uniformity of the color within a given patch verifies that all the alpha plates within it have the same orientation as expected. In addition, the irregular contours of the patches bear a resemblance to the irregular outlines of the colonies thus verifying that they are indeed representative of these microstructural features rather than the prior-beta grains whose shapes are largely equiaxed. In contrast, the BAWQ microtexture showed less distinct patches of color (Figure 11c, d). This can be ascribed to the fact that each colony (or beta grain) contains several different variants of acicular plates. Despite the fact that the as-heat treated BAWQ material did not appear to have a noticeable microtexture, the fact that it exhibited a strong macrotexture (Figure 7b, Table 2) indicates that relatively few variants were selected during the beta-to-alpha prime transformation.

After hot rolling, the BACC material retained evidence of microtexture. Examples from samples rolled to a large reduction at 815 or 925°C are shown in Figure 12. In both cases, the color patches have an elongated shape. Bearing in mind that the EBSP measurements were taken on cross sections containing the transverse direction and sheet normal, this elongation is as expected for the plane-strain reduction characterizing rolling. To be specific, after a 77 percent reduction (i.e., the deformation in the heavy reduction experiments), a colony measuring originally 250 x 250 μm in a transverse cross section would deform to a size of approximately 60 x 250 μm , in rough agreement with the observations. The plots in Figure 12 also reveal that the deformed colonies exhibit relatively few different colors. This behavior can be explained on the basis of the development of a sharp macrotexture (including the disappearance of the 45° macrotexture component) which would lead to a reduction in

the overall spread of orientations and thus number of different colors that would be evident on icosahedral or Eulerian plots.

Taken together, the microstructure, macrotexture, and microtexture data for the hot rolled BACC Ti-6Al-4V material provide insight into the globularization process during hot working. The results suggest large macrotexture changes at low strains due to plate bending/kinking but with limited globularization of the plates. Because a microtexture is retained to high strains, it can be inferred that discontinuous recrystallization (which would tend to randomize the orientations within a given colony) is probably not occurring to a significant extent. Rather, strain localization processes within the alpha plates and subsequent penetration of the beta phase, as described by Weiss, et al. [3,21], can be deduced to be the mechanism by which globularization would proceed while maintaining the similar identity of globularized alpha grains within a prior colony.

Post-rolling heat treatment of BACC samples resulted in a significant change in microtexture, but not necessarily a reduction in its prominence. For example, the evolution of microtexture for a BACC sample that was lightly rolled at 925°C and subsequently heat treated at 925°C is shown in Figure 13. There was not a noticeable reduction in size.

Microtexture developed in a BAWQ sample via rolling to a heavy reduction at 970°C is shown in Figure 14. Similar to the as-heat treated condition (Figure 11c, d), the rolled BAWQ samples did not exhibit a very strong microtexture. Furthermore, as for the heavily rolled BACC samples, the microtexture plots for the rolled BAWQ

samples revealed few colors indicating a narrow spread in orientations, a trend that would be expected based on the development of a strong macrotexture.

4. Conclusions

The evolution of microstructure, macrotexture, and microtexture during the thermomechanical processing of Ti-6Al-4V with transformed beta starting microstructures was determined. The following conclusions are drawn from this work:

1. Moderately high reductions (~ 75 pct.) are required to globularize colony or acicular (martensitic alpha) microstructures dynamically via hot rolling. Lower reductions (~ 40 pct.) in conjunction with near subtransus post-rolling heat treatments may also be used to effect globularization.
2. Beta annealing followed by controlled cooling (to form a lamellar colony structure) or water quenching (to form martensitic alpha) does not produce a random macrotexture.
3. The principal alpha macrotexture component developed during unidirectional hot rolling consists of one for which the basal poles are aligned transverse to the rolling direction. The predominance of this TD texture component even at low reductions suggests that kinking and bending of lamellar plates at low strains play a key role in the development of macrotexture and the occurrence of flow softening.
4. The persistence of noticeable microtexture during the subtransus hot rolling of Ti-6Al-4V with a lamellar colony starting microstructure suggests that globularization of

the alpha plates does not occur by recrystallization but by localized deformation of the plates and penetration by the beta phase.

5. Strong macrotextures in Ti-6Al-4V with transformed beta or globularized alpha microstructures can be associated with either weak or strong microtextures depending on the nature of the transformed beta (i.e., lamellar colony versus basketweave/acicular).

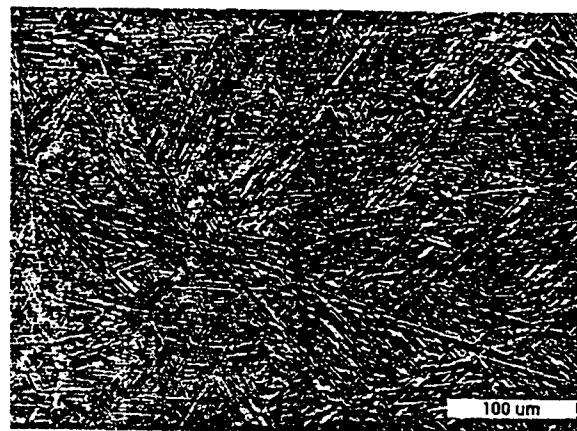
References

1. S. L. Semiatin, V. Seetharaman, and I. Weiss, in I. Weiss, R. Srinivasan, P. J. Bania, D. Eylon, and S. L. Semiatin (eds.), *Advances in the Science and Technology of Titanium Alloy Processing*, TMS, Warrendale, PA, 1997, pp. 3-73.
2. S. L. Semiatin, J. F. Thomas, Jr., and P. Dadras, *Metall. Trans. A*, **14A** (1983) 2363-2374.
3. I. Weiss, F.H. Froes, D. Eylon, and G. E. Welsch, *Metall. Trans. A*, **17A** (1986) 1935-1947.
4. E. Rauch, G.R. Canova, and J.J. Jonas, McGill University, Montreal, Canada, unpublished research, 1982.
5. J. G. Malcor, F. Montheillet, and B. Champin, in G. Luetjering, U. Zwicker, and W. Bunk (eds.), *Titanium: Science and Technology*, Deutsche Gesellschaft für Metallkunde e.V., Oberursel, Germany, 1985, pp. 1495-1502.
6. A. A. Korshunov, F. U. Enikeev, M.I. Mazurskii, G. A. Salishchev, A.V. Muravlev, P. V. Chistyakov, and O.O. Dimitriev, *Russ. Metall.*, **3** (May - June 1994) 103-108.
7. M. Brun and G. Shachanova, in I. Weiss, R. Srinivasan, P. J. Bania, D. Eylon, and S.L. Semiatin (eds.), *Advances in the Science and Technology of Titanium Alloy Processing*, TMS, Warrendale, PA, 1997, pp. 193-199.
8. F. Larson and A. Zarkades, *Properties of Textured Titanium Alloys*, Report MCIC-74-20, Metals and Ceramics Information Center, Battelle's Columbus Laboratories, Columbus, OH, 1974.
9. M. F. Amateau, D.L. Dull, and L. Raymond, *Metall. Trans.*, **5** (1974) 561-564.
10. G. Luetjering and A. Gysler, in G. Luetjering, U. Zwicker, and W. Bunk (eds.), *Titanium: Science and Technology*, Deutsche Gesellschaft für Metallkunde e.V., Oberursel, Germany, 1985, pp. 2065-2083.
11. A. P. Woodfield, M. D. Gorman, R.R. Corderman, J. A. Sutliff, and B. Yamron, in P. A. Blenkinsop, W. J. Evans, and H.M. Flower (eds.), *Titanium '95: Science and Technology*, Institute of Metals, London, 1996, pp. 1116-1123.
12. J. S. Kallend, U.F. Kocks, A.D. Rollett, and H.R. Wenk, *Mater. Sci. Eng. A*, **A152** (1992) 9-18.

13. B.L. Adams, S.I. Wright, and K. Kunze, *Metall. Trans. A*, **24A** (1993) 819-831.
14. T.A. Mason and B.L. Adams, *JOM*, **46** (October 1994) 43-45.
15. S.L. Semiatin and V. Seetharaman, Materials and Manufacturing Directorate, Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, unpublished research, 1997.
16. H. Moustahfid, N. Gey, M. Humbert, and M.J. Philippe, *Metall. and Mater. Trans. A*, **28A**, (1997) 51-61.
17. M. Peters and G. Luetjering, in H. Kimura and O. Izumi (eds.), *Titanium '80: Science and Technology*, TMS, Warrendale, PA, 1980, pp. 925-935.
18. N. Gey, M. Humbert, M.J. Philippe, and Y. Combres, *Mater. Sci. Eng. A*, **A219** (1996) 80-88.



(a)



(b)

Figure 1. Optical microstructures of Ti-6Al-4V program materials: (a) beta annealed and controlled cooled (BACC), (b) beta annealed and water quenched (BAWQ).

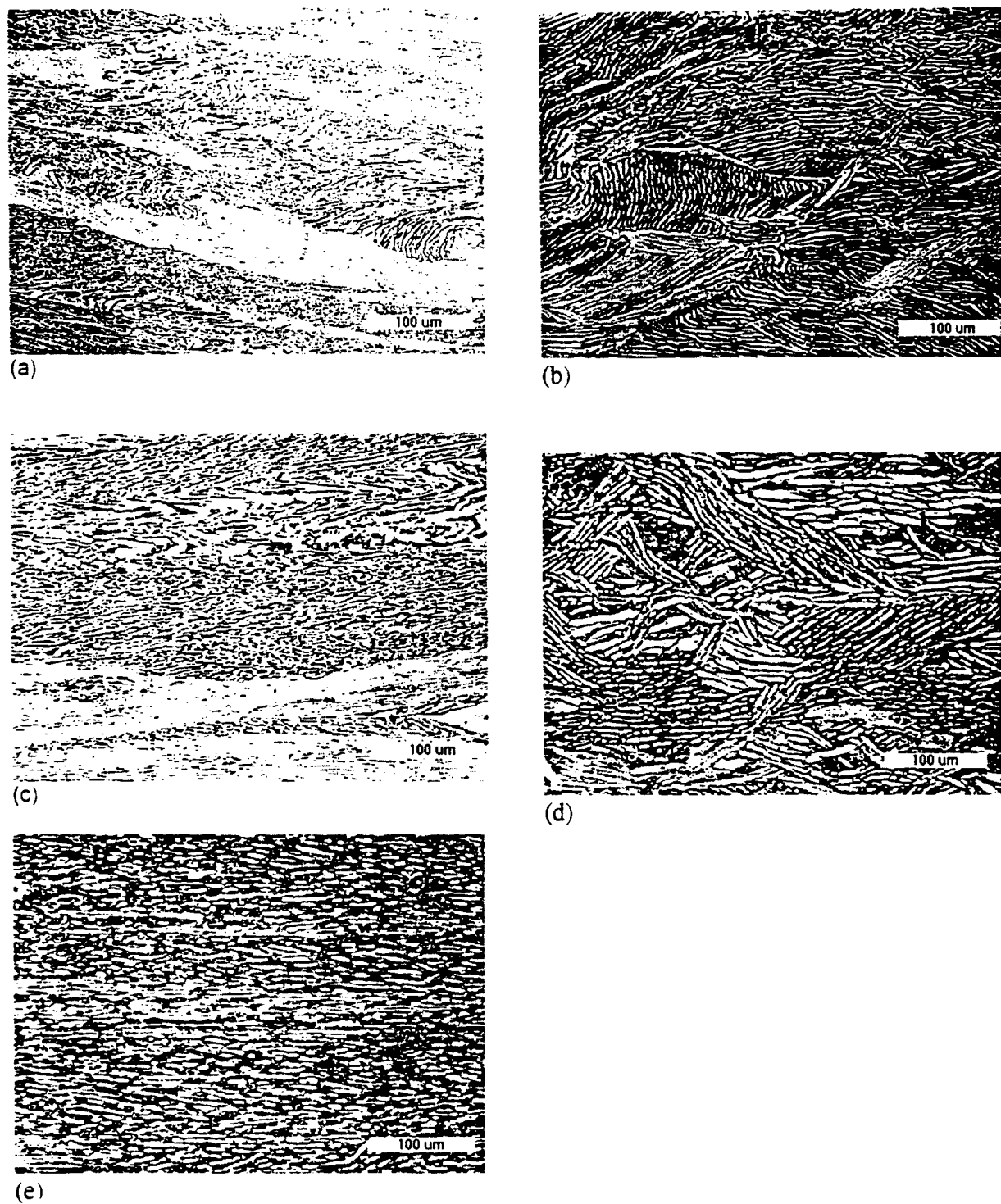


Figure 2. Optical microstructures developed in hot rolled Ti-6Al-4V BACC samples. Rolling temperatures and total reductions (in percent) were: (a) 815°C/77, (b) 925°C/44, (c) 925°C/77, (d) 970°C/44, and (e) 970°C/77. In all micrographs, the rolling direction is horizontal, and the thickness direction is vertical.

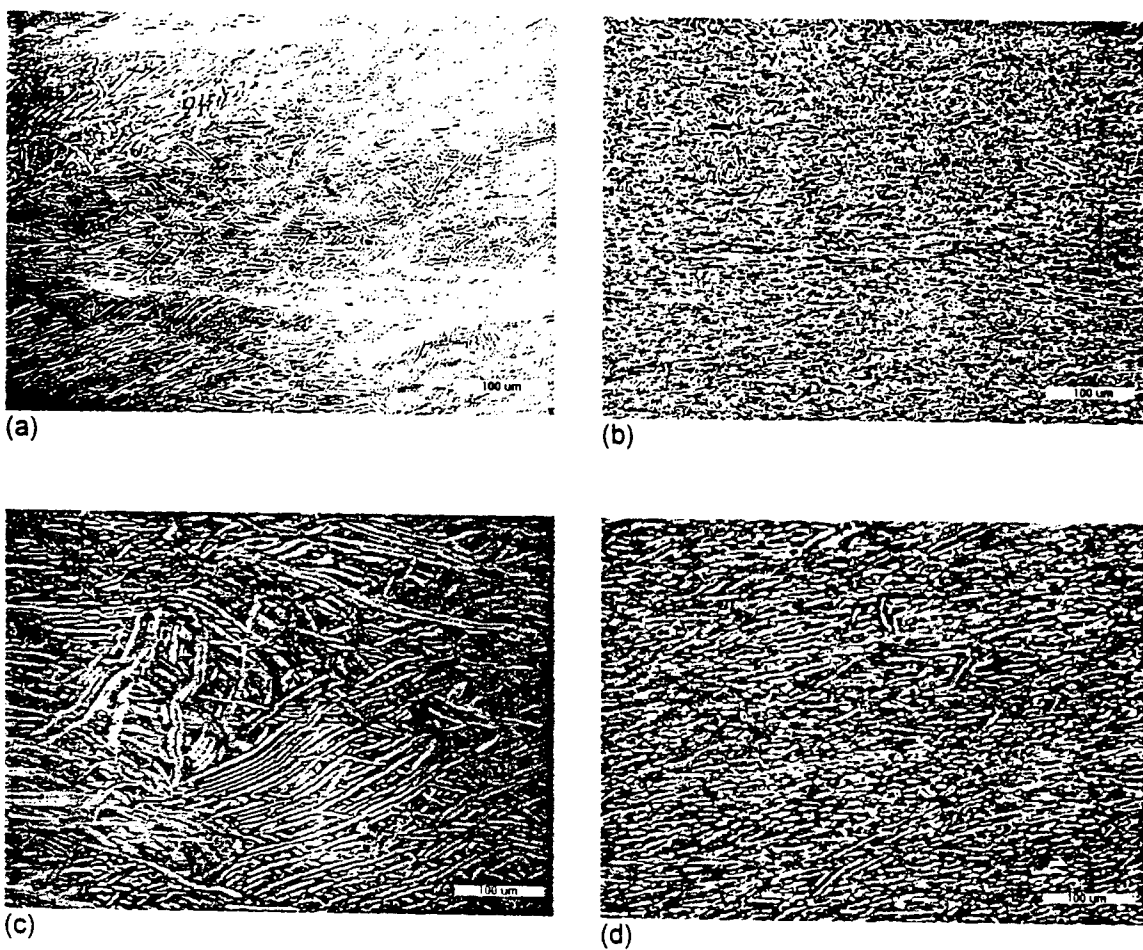


Figure 3. Optical microstructures developed in hot rolled Ti-6Al-4V BAWQ samples. Rolling temperatures and total reductions (in percent) were: (a) 925°C/44, (b) 925°C/77, (c) 970°C/44, and (d) 970°C/77. In all micrographs, the rolling direction is horizontal, and the thickness direction is vertical.

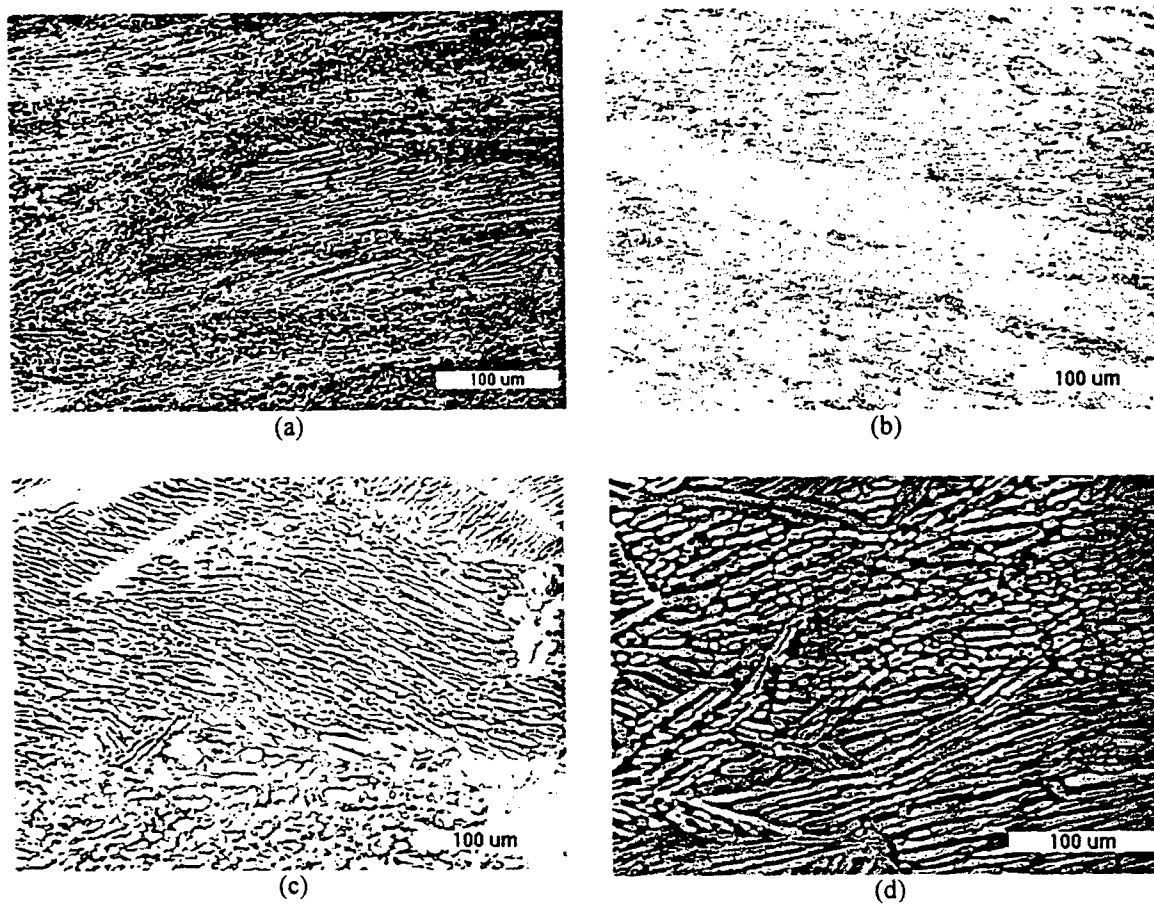


Figure 4. Optical microstructures developed in hot rolled Ti-6Al-4V BACC samples, after solution annealing at 870°C. Rolling temperatures and total reductions (%) were: (a) 815°C/77, (b) 925°C/77, (c) 925°C/44, and (d) 970°C/44. In all micrographs, the rolling direction is horizontal, and the thickness direction is vertical.

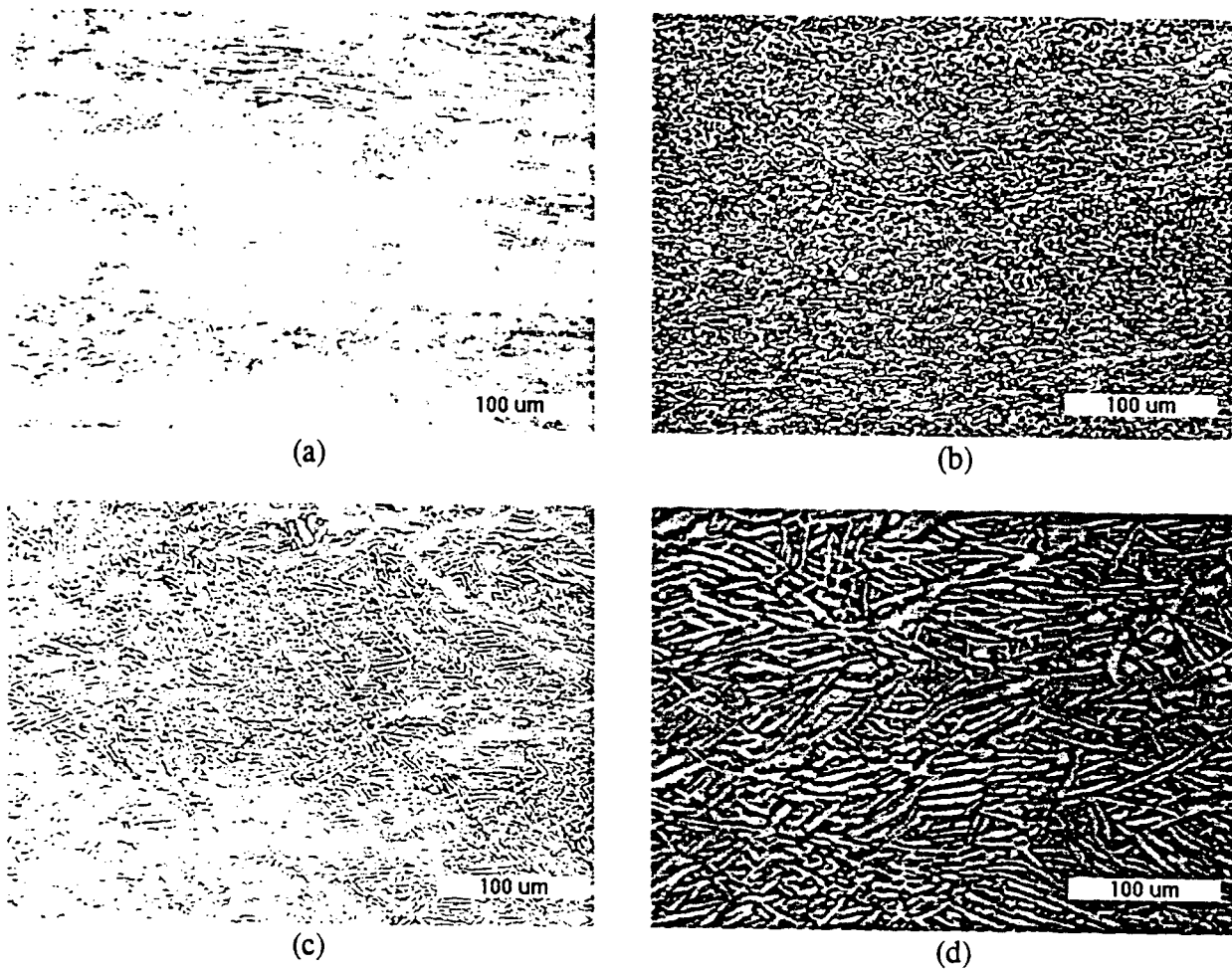


Figure 5. Optical microstructures developed in hot rolled Ti-6Al-4V BAWQ samples, after solution annealing at 870°C. Rolling temperatures and total reductions (%) were: (a) 970°C/77, (b) 925°C/77, (c) 925°C/44, and (d) 970°C/44. In all micrographs, the rolling direction is horizontal, and the thickness direction is vertical.

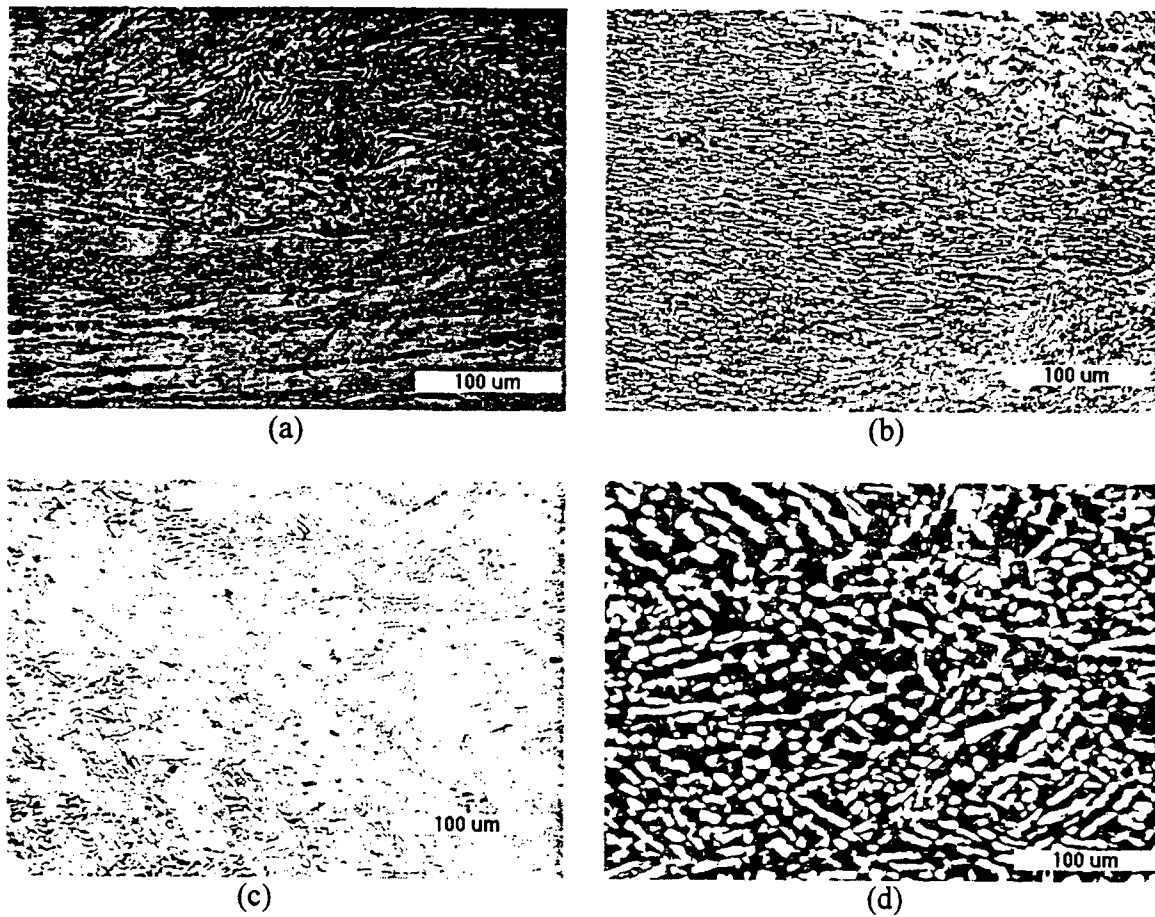


Figure 6. Optical microstructures developed in hot rolled Ti-6Al-4V samples, after solution annealing at a the same temperature as their rolling temperature. Preform microstructure/rolling temperature/total reduction (in percent)/heat treatment temperature were as follows: (a) BACC/815°C/77/815°C, (b) BACC/925°C/77/925°C, (c) BAWQ/925°C/44/925°C, (d) BAWQ/970°C/44/970°C. In all micrographs, the rolling direction is horizontal, and the thickness direction is vertical.

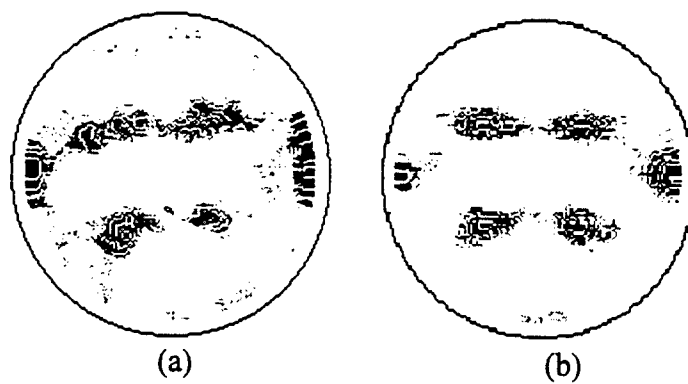


Figure 7. (0002) alpha-phase pole figures for Ti-6Al-4V rolling preform material:
(a) BACC condition, (b) BAWQ condition.

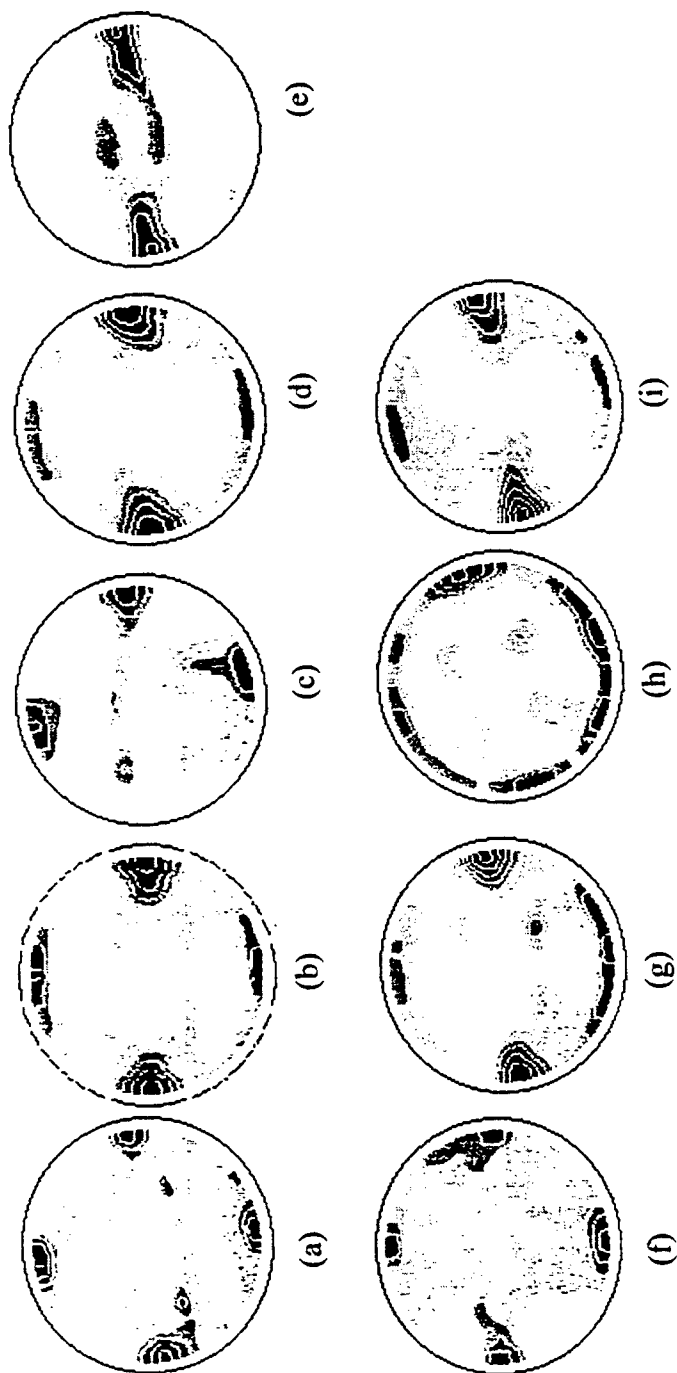


Figure 8. (0002) alpha-phase pole figures for hot rolled Ti-6Al-4V preforms. Preform microstructure/rolling temperature/total reduction (in percent) were as follows: (a) BACC/970°C/44, (b) BACC/970°C/77, (c) BACC/925°C/44, (d) BACC/925°C/77, (e) BACC/815°C/77, (f) BAWQ/970°C/44, (g) BAWQ/970°C/77, (h) BAWQ/925°C/44, (i) BAWQ/925°C/77

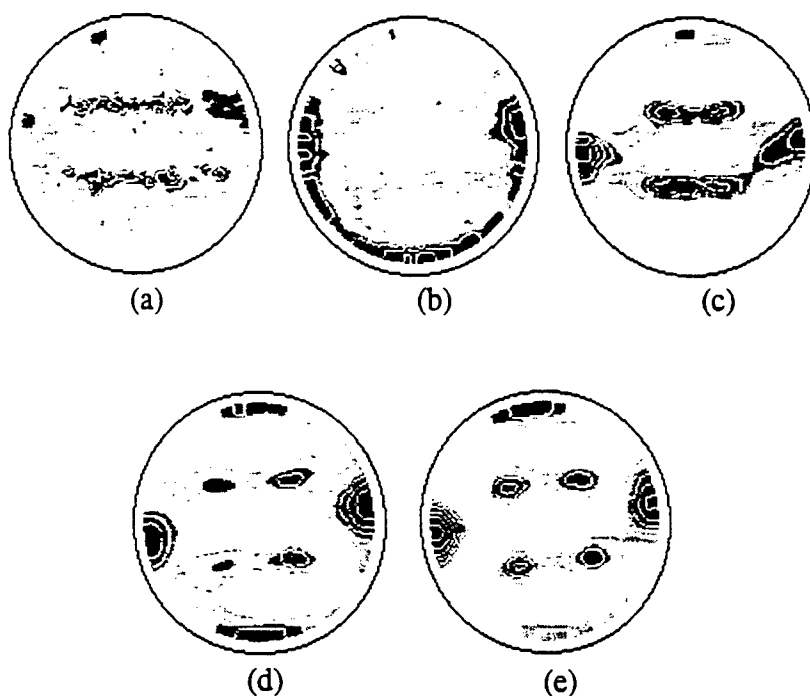


Figure 9. (0002) alpha-phase pole figures for hot rolled Ti-6Al-4V samples. Preform microstructure/rolling temperature/total reduction (%) /heat treatment temperature were as follows:
 (a) BACC/970°C/77/955°C, (b) BACC/925°C/77/955°C, (c) BACC/815°C/77/955°C, (d) BAWQ/970°C/77/955°C, (e) BAWQ/925°C/77/955°C

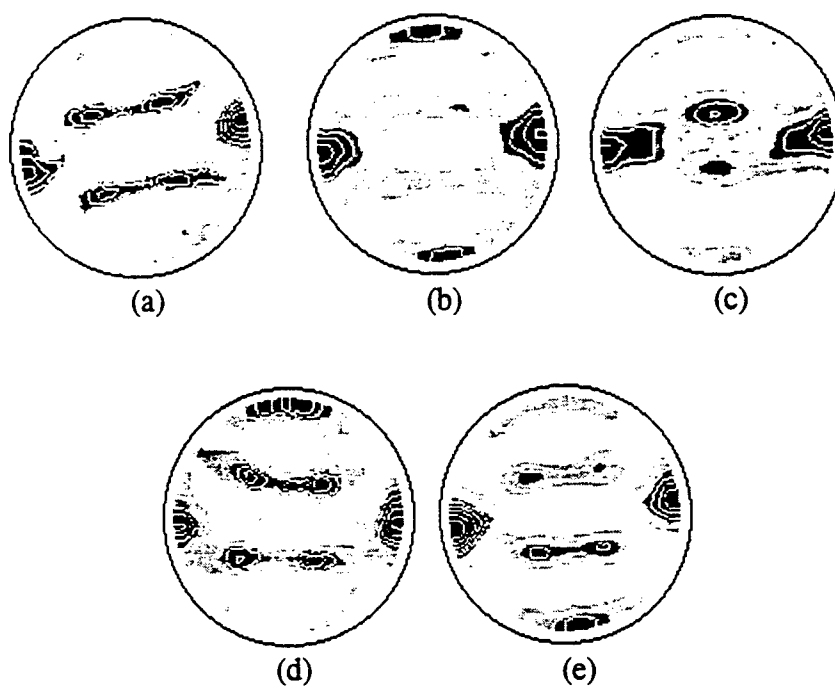


Figure 10. (0002) alpha-phase pole figures for hot rolled Ti-6Al-4V samples. Preform microstructure/rolling temperature/total reduction (in percent)/heat treatment temperature were as follows: (a) BACC/970°C/77/970°C, (b) BACC/925°C/77/925°C, (c) BACC/815°C/77/815°C, (d) BAWQ/970°C/77/970°C, (e) BAWQ/925°C/77/925°C.

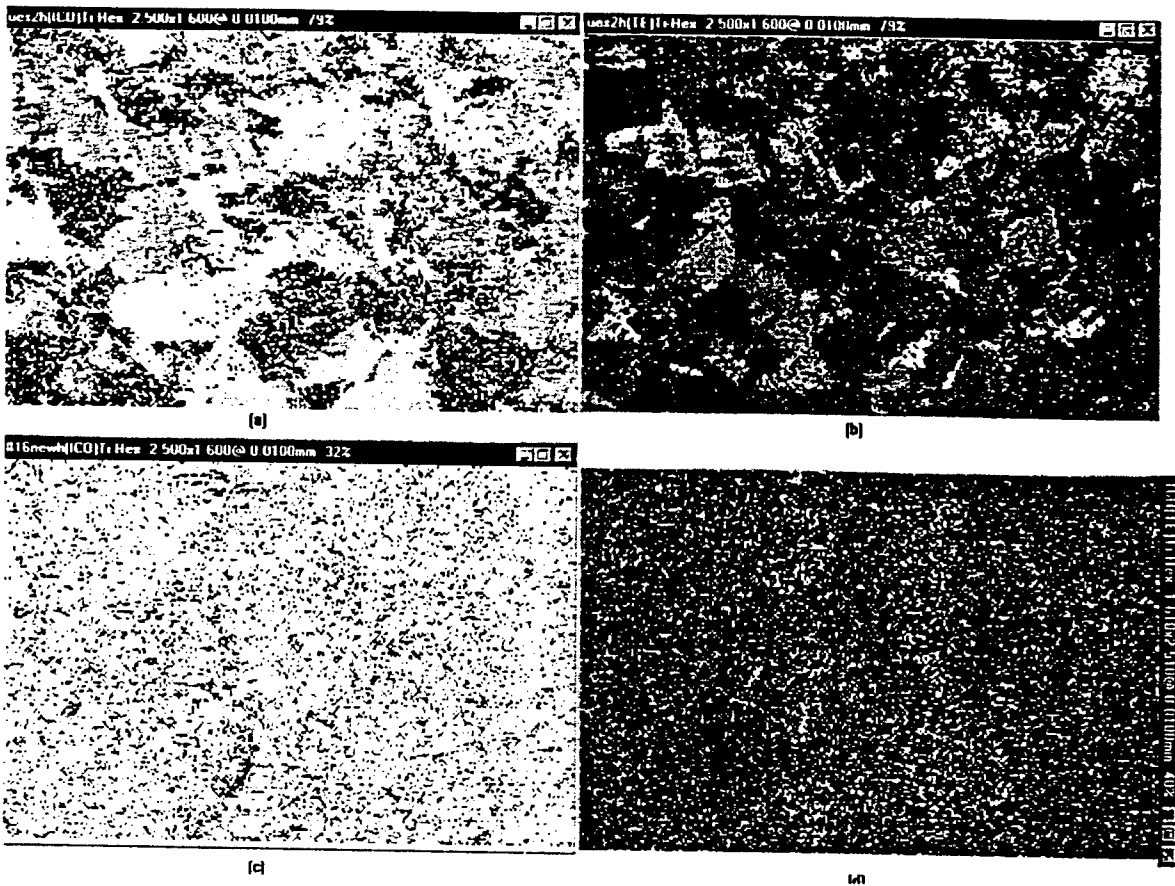
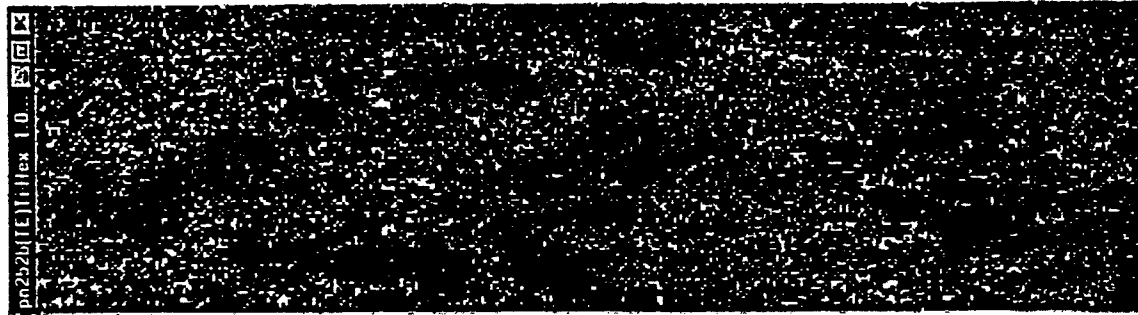
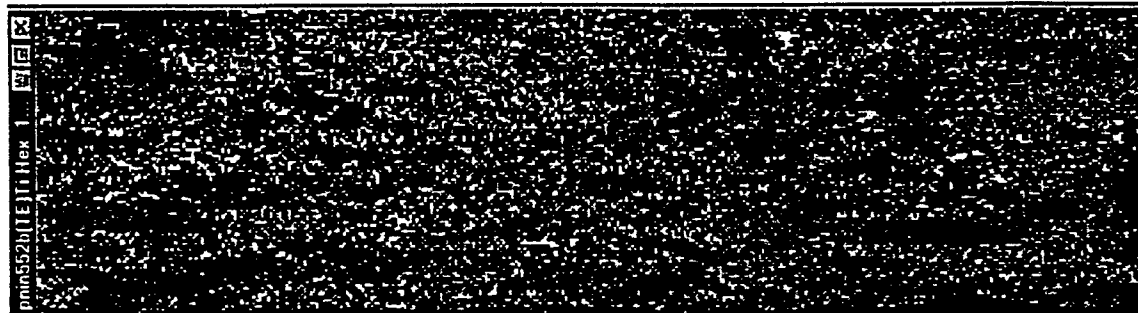


Figure 11. Initial (mid plane) alpha-phase microtextures for Ti-6Al-4V rolling preforms: (a,b) BACC condition, and (c,d) BAWQ condition. (a,c) are icosahedral plots and (b,d) are total Eulerian plots.



(a)



(b)

Figure 12. Midplane alpha-phase microtextures for Ti-6Al-4V BACC samples rolled to 77 percent reduction in thickness at (a) 815°C or (b) 925°C. Both plots are total Eulerian representations.

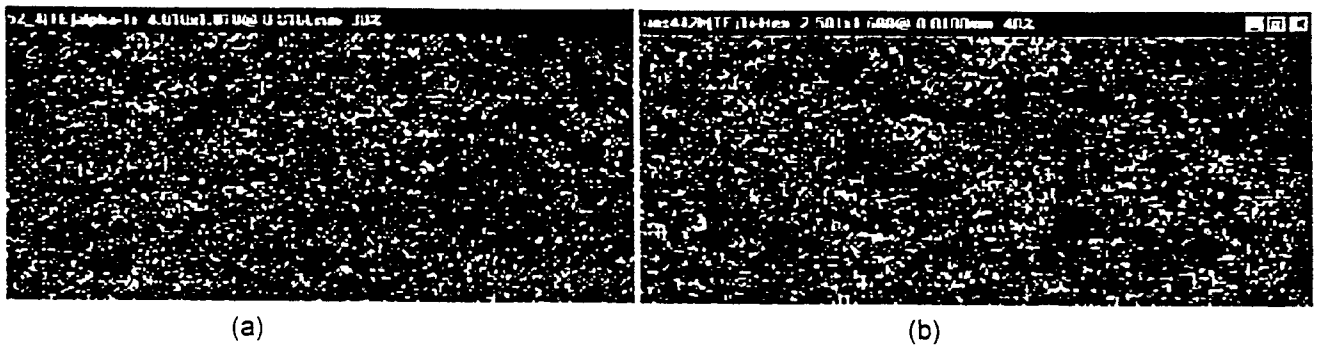


Figure 13. Microtextures of a BACC sample that was rolled to a 44 percent reduction in thickness at 925°C: (a) as hot rolled, or (b) rolled and heat treated at 925°C.

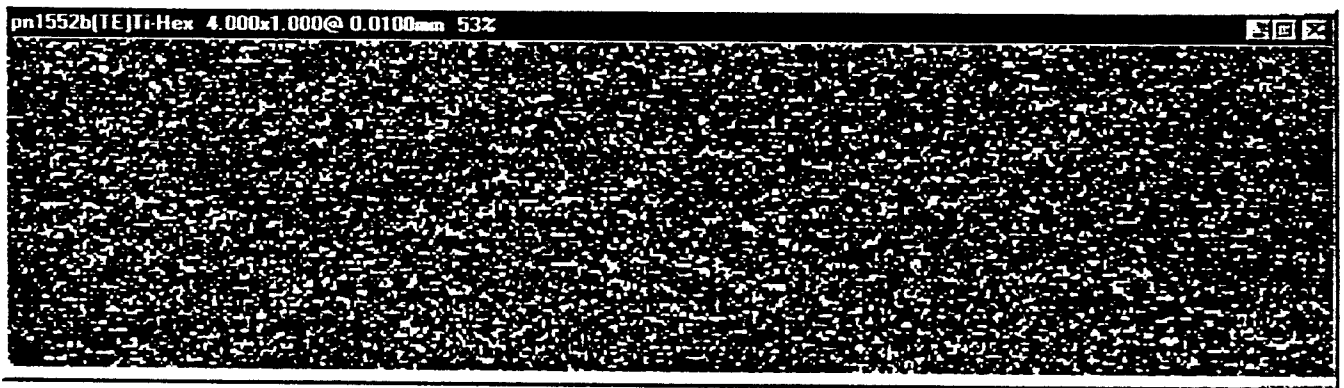


Figure 14. Total Eulerian representation of the midplane alpha-phase microtexture for a Ti-6Al-4V BAWQ sample rolled to 77 percent reduction in thickness at 970°C.

MULTI-SITE AND WIDESPREAD FATIGUE DAMAGE IN AIRCRAFT STRUCTURE IN
THE PRESENCE OF PRIOR CORROSION

J.D. Baldwin
Assistant Professor
School of Aerospace & Mechanical Engineering

University of Oklahoma
865 Asp Avenue, Room 212
Norman, Oklahoma 73019

Final Report For
Summer Research Extension Program
Wright Laboratory, Wright-Patterson AFB, Ohio

Sponsored By:
Air Force Office of Scientific Research
Bolling Air Force Base, D.C.

January 30, 1998

MULTI-SITE AND WIDESPREAD FATIGUE DAMAGE IN AIRCRAFT STRUCTURE IN THE PRESENCE OF PRIOR CORROSION

J.D. Baldwin
Assistant Professor
School of Aerospace & Mechanical Engineering
University of Oklahoma

Abstract

This study explores the impact of corrosion-induced pillowing on stress intensity factors for corner cracks in countersunk rivet holes. Such deformation patterns have been observed in aircraft fuselage lap joints. The boundary element method was used to compute the stresses and stress intensity factors for a section of lap joint loaded by faying surface pressure as well as the fuselage pressurization hoop stress. Circular-front corner cracks were constructed in the countersink varying in length from 0.01 to 0.08 inches at angular orientations varying from 10 degrees to 80 degrees from the hoop stress direction. The stress analysis showed that a distinct compressive region is produced in the countersink by the pillowing pressure and that this compression is not cancelled out by the hoop stress. Such a compressive region is felt to be sufficient to cause any cracks that form in the countersink to be driven toward the joint faying surface.

MULTI-SITE AND WIDESPREAD FATIGUE DAMAGE IN AIRCRAFT STRUCTURE IN THE PRESENCE OF PRIOR CORROSION

J.D. Baldwin

Introduction and Background

As a result of military budget constraints and the quest for greater commercial competitiveness, aircraft are now being utilized far beyond their original designed operational life spans. The United States Air Force operates over 600 Boeing C/KC-135 aircraft that have been in service 35 years or more. These aircraft were designed in the 1950's and are cousins to the civilian Boeing 707. Because these aircraft are crucial elements in the force projection strategy of the USAF, it has been proposed that they should remain in service well past the year 2017.¹

Corrosion was an unforeseen factor not included in the initial service life extension studies. However, it is pervasive, unpredictable and often undetectable. Corrosion damage repair costs the USAF more money than any other structurally related item. A recent study funded by Warner Robins Air Logistics Center conservatively estimated that corrosion alone costs the USAF \$700 million annually.² Since the costs of replacing this fleet of large jet aircraft are beyond the scope of current budgets, Boeing and the USAF have been forced to confirm the feasibility of extending their operational lives.

Figure 1 schematically illustrates the criteria for determining inspection schedules and the potential effects of corrosion damage on these schedules. In the damage tolerance approach used by the USAF, a crack of length a_0 is assumed to exist in the structure at the time of its manufacture. This length is generally considered to be the maximum length that is undetectable

by current inspection procedures. The crack is further assumed to exist at the worst-case location and orientation. The non-corroded a versus N curve is then used to predict the number of loading cycles, N_1 , necessary for the crack to grow from its initial length, a_0 , to the critical length at which structural failure would occur, a_c . Inspections would then be scheduled at half this number of cycles, $\frac{1}{2} N_1$, to identify a crack of length a_i , triggering repair or replacement of the damaged structure. Inspections are scheduled at half of the critical number of cycles in order to afford two inspections before the crack reaches critical length. The effects of corrosion, though, could alter these predictions in two ways. Material damage and changes in the stress state may tend to decrease the critical crack length and accelerate the rate of crack growth. These effects are seen as a change in a_c and a steeper a versus N curve in Figure 1. As a result, the number of cycles to reach the critical crack length is reduced to N_2 . If corrosion is neglected and the inspection schedule is not changed to $\frac{1}{2} N_2$, the likelihood of detecting a crack before it reaches its critical length is reduced.

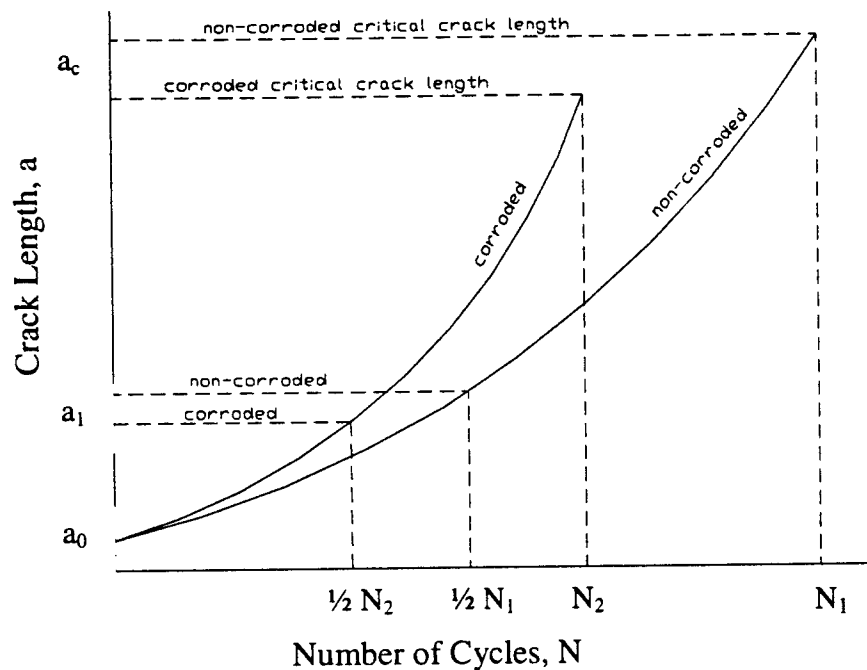


Figure 1: Crack length vs. number of fatigue cycles indicating critical points and the effects of corrosion on crack growth.

Initial service life extension estimates for the C/KC-135 fleet have been based on fatigue damage accumulation alone, and predict that the fleet will be viable until at least the year 2020. However, as a result of decades of exposure to both natural elements and pollution, corrosion has become an increasingly important factor in the deterioration of airframe structures. This may lead to a reduction in the inspection intervals needed to detect damage and maintain a safe fleet, as illustrated in Figure 1. Quantification of the impact of corrosion on crack growth and residual strength is still a topic of great interest. The experimentally determined material response data do not yet exist to form hard conclusions regarding the modification of inspection intervals.

For the damage tolerance philosophy to be truly successful, it is important to determine the level of corrosion damage at which fatigue life is seriously affected. The damaging effects of corrosion manifest themselves in a variety of ways. These include surface pitting, exfoliation, stress corrosion cracking, and pillowing. Isolated pitting, in and of itself, has little effect on the

strength of a structure. It can, though, result in a multitude of crack initiation sites that could lead to fatigue damage. More widespread pitting can result in an effective reduction of material thickness, limiting the load carrying capacity of the structure. Exfoliation is caused by cracks propagating parallel to material surfaces resulting in a significant reduction in the effective material thickness. Stress corrosion cracking is most often an intergranular separation, the mechanisms of which are still under debate. Pillowing is a corrosion phenomenon that is particularly common in fuselage lap joints. As the faying surfaces between the lap joint layers corrode, the trapped corrosion products build up leading to out-of-plane deformations (Figure 2). The USAF aircraft maintenance community is keenly interested in the effects of pillowing on fatigue cracking.

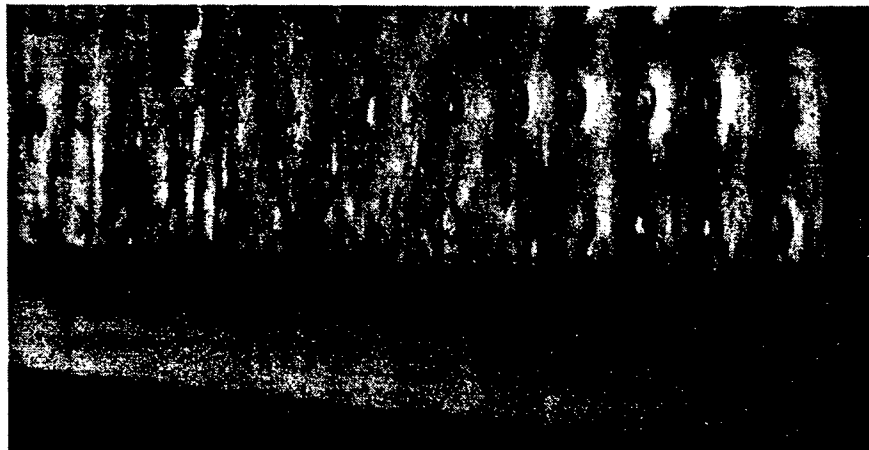


Figure 2: D-Sight image of a simple shear lap joint. Corrosion pillowing is visible to the right.³

Sites of particular interest are countersunk rivet holes in pillowed fuselage joints. The processes of machining rivet holes and inserting fasteners often damage aircraft skins sufficiently to create possible crack initiation sites. The combination of these material imperfections with the stress concentrations inherent to holes results in ideal conditions for the initiation and propagation of corner cracks. A corner crack is a specific form of surface crack characterized by

the intersection of the crack with both surfaces that intersect to form a corner. Surface cracks do not extend entirely through the material thickness, which distinguishes them from through cracks. One troubling phenomenon that has been observed in pillowed lap joints on operational aircraft is the tendency for surface cracks to initiate on the faying surface, possibly as corner cracks, and propagate into high aspect ratio elliptical cracks (Figure 3) without breaking through the top surface. This results in very large cracks that can not be detected by visual inspection. Another characteristic of these cracks is their orientation away from normal to the hoop stress (Figure 4).

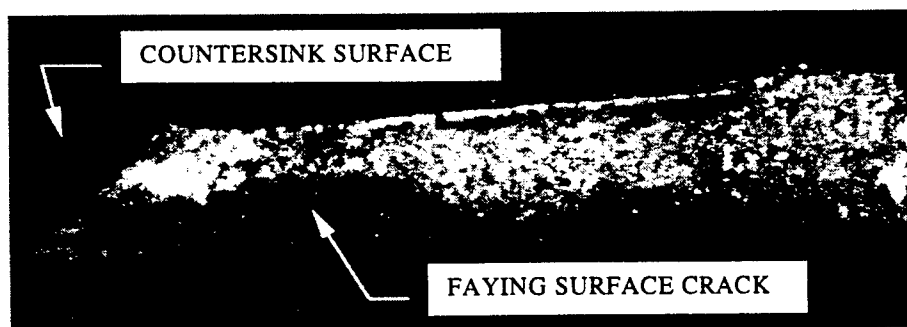


Figure 3: Cross-section of an upper lap joint layer showing high aspect ratio crack along the faying surface.³

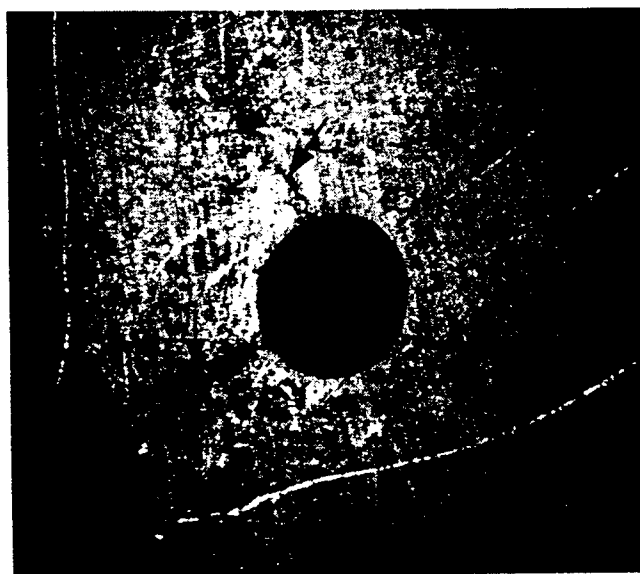


Figure 4: Faying surface cracks not aligned with the normal to the hoop stress (cracks indicated by arrows). The crack in Figure 3 appears at the lower right side of the rivet hole.³

Previous Research

Recent efforts have been made to further our understanding of the stress states and altered material properties that result from corrosion induced pillowing deformation. In the past, stress and fatigue analysts used a reduction in the nominal material thickness to account for the effects of corrosion. This, however, severely underestimates the damage caused by corrosion by failing to take into account the significantly altered stress state that results from pillowing. Studies investigating the stresses created by corrosion and the resultant pillowing have been conducted by Komorowski,⁴ et al., at the National Research Council - Canada. In their finite element studies, a uniform pressure applied to the faying surface was used to simulate the forces resulting from a buildup of corrosion products. Their results have indicated that if faying surface corrosion exists, the stress increase due to thickness reduction is negligible when compared to the stresses induced by pillowing. This holds true for thickness reduction and pillowing compared separately and in combination. Therefore, modeling pillowing alone provides the best combination of realism and simplicity. Another result of the research by Komorowski, et al., was a correlation between the level of corrosion, measured as a percent thickness loss, and the maximum out-of-plane displacement caused by a buildup of corrosion products. The investigators were attempting to obtain calibration data to be applied to D-Sight (Diffracto Ltd., Windsor, Ontario, Canada) optical nondestructive inspection (NDI) techniques. D-Sight is a surface inspection technique that uses a light source and retro-reflective screen. This could allow the quick and accurate determination of the extent of corrosion in lap joints while avoiding many of the problems associated with other NDI methods.⁵ Their calculations were based on the observation that typical corrosion products in lap joints consist mainly of aluminum oxide

trihydrate and a proportionally small amount of aluminum oxide monohydrate. A volume ratio for aluminum oxide trihydrate to pure aluminum of 6.5 to 1 was used to calculate the volume required to accommodate the products generated by corrosion of a given percentage of material thickness. Using a uniform pressure on the lower surface of the outer skin, the deflections necessary to contain the corrosion products were calculated using the closed-form classical plate theory.⁶ Only the upper skin was modeled due to the assumed symmetry of the deformed joint.

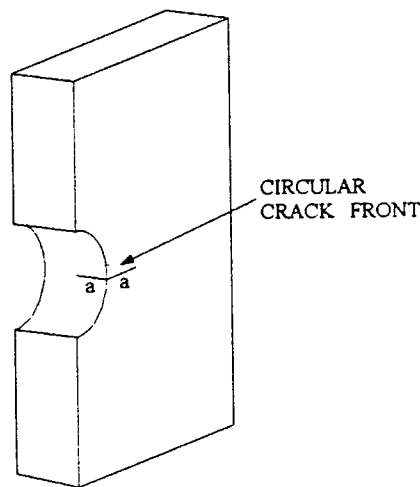


Figure 5: Geometry of a round corner crack in a straight bore hole.

Because the cracks in the countersunk hole may begin as corner cracks, they can be thought of as extensions to the case of round corner cracks in a straight bore hole. The geometric definition of this crack case is shown in Figure 5. The first detailed examination of the stress intensity factors in corner cracks in straight bore holes was given by Smith and Kullgren⁷ in 1977 by means of the finite element alternating method. In 1979, Raju and Newman⁸ used a three-dimensional finite element method to compute stress intensity factors for two symmetric corner cracks on opposite sides of a straight bore hole. Analysis of these data was subsequently presented in the form of curve fits⁹ that permitted interpolation between the analysis cases.

Nishioka and Atluri¹⁰ solved the same symmetric crack cases considered by Raju and Newman⁸ using the finite element alternating method and found good agreement with the prior results except at the bore of the hole where the two solutions diverged. The source of the error in the earlier finite element solutions was subsequently found to be poorly shaped elements at the hole surface.¹¹ Revised finite element solutions were given by Shivakumar and Newman¹² in 1991. In 1990, Zhao and Wu¹³ presented a weight function solution for the corner crack problem by combining a series of two-dimensional solutions to form the three-dimensional stress intensity data. It is important to note that all of the prior solutions referenced above have been for straight bore holes (i.e., with no countersink feature). The only known set of solutions for stress intensity factors in countersunk holes is due to Tan, et al.¹⁴. However, these studies have focused primarily on uniaxial tension and the pure fatigue crack growth resulting from variable loading. The unique cracking conditions that exist near rivet holes in corroded lap joints have been neglected. It was the goal of the research presented here to make a first attempt at describing the effects of pillowing on the properties of elliptical cracks in countersunk rivet holes.

Objectives

Unlike the simple two-dimensional through crack, real cracks exist in three dimensions, are often elliptical in shape, and do not extend entirely through the body. Under these conditions, the evaluation of stress intensity becomes a more complex matter, but it is still a function of the stress field and crack geometry. Various numerical techniques based on crack face displacements are employed to facilitate these calculations,¹⁵ but will not be described in detail as they are not essential to the concept of stress intensity factors. One goal of this study was to provide a preliminary analytical database of Mode I stress intensity factors for cracks emanating

from countersunk rivet holes at different angular orientations, with the focus on the geometry and conditions found in C/KC-135 fuselages. The second objective was to model three-dimensional crack propagation into the pillowed regions and to identify the differences between these cracks and the planar cracks used in the preliminary investigation of stress intensity factors. The remainder of this report will discuss the tools and modeling techniques used to achieve these goals, as well as the results of the study.

Methods of Analysis

The tool chosen for the calculation of stress intensity factors was *FRANC3D*, a boundary element program developed by the Cornell University Fracture Group lead by Dr. Anthony Ingraffea.¹⁶ Unlike finite element methods (FEM) which require the discretization of the body volume, boundary element methods (BEM) require only the discretization of the body surface. This inherently leads to fewer elements being needed to accurately represent a body, especially for three-dimensional problems. Since the complexities of countersunk hole geometry with both in-plane and out-of-plane loading were such that a three-dimensional analysis was necessary, BEM offered advantages where model generation efforts were concerned. Another benefit of requiring fewer elements to describe the structure is a decrease in the number of algebraic equations that must be solved. In addition, BEM reduces the dimensionality of any given problem by one. For instance, a three-dimensional problem results in a two-dimensional boundary integral equation. However, because boundary element methods produce fully populated system matrices for homogeneous regions, BEM cannot use the efficient sparse matrix solvers employed in the reduction of the larger, sparsely populated finite element matrices. Thus, the boundary element method requires a greater number of computations to obtain each

component of the solution matrix compared to FEM. This offsets to some degree the computational advantages of fewer equations, lower dimensionality, and decreased matrix reduction. The benefits of BEM are best realized when working with models of large physical dimensions.¹⁷

The major advantage of *FRANC3D* and its supporting software was their ability to facilitate solid modeling and discretization of complex geometries. *FRANC3D* uses a hierarchy of models that combine to fully describe the geometry, material properties, and boundary conditions of the structure to be analyzed. The basis of the hierarchy is the solid model. Unlike most engineering simulations in which the structure and boundary conditions must be described entirely by a mesh and its attributes, *FRANC3D* uses the solid model to specify geometry, material properties, and boundary conditions. These properties are, therefore, part of the basic description of the structure and independent of any mesh applied for analytical purposes. Initial uncracked geometries are created using the Object Solid Modeler (*OSM*) software developed by the Cornell Fracture Group. These geometry files are then interpreted by *FRANC3D* where boundary conditions and material properties are applied directly to geometric volumes, faces, edges, and vertices. Material properties are assigned to the three-dimensional continuum and become a part of the basic model. The analytical model, or mesh, is then overlaid on the solid model and inherits the properties assigned to the solid model. In this way, different meshes can be applied to the model without the user needing to redefine the elemental or nodal attributes. The stress analysis is done by the Boundary Element System (*BES*) software. *FRANC3D* provides post-processing functions. A significant limitation of *FRANC3D* (and most BEM systems) is that it is limited to linear, elastic analysis, which dictated that the stresses induced in the model must remain below the yield point of the material.¹⁹

Cracks in *FRANC3D* are created in the geometry model and are, therefore, inherited by the analytical model. The capability exists to model internal cracks, through cracks, and surface cracks, though only surface cracks were considered in this study. Surface cracks have at least one intersection with a boundary surface. Corner cracks like the ones studied here intersect at least two surfaces (Figure 5) and possibly more, depending on the particular crack geometry. Cracks are modeled as a main face and a mate face that are topologically separate but geometrically identical. This also applies to the edges and vertices that define the crack faces, with the exception of the crack front edges and vertices. This allows the crack faces to separate in the analytical model without imposing an artificial finite gap in the geometry. Such a gap could cause computational errors introduced by rounding and point co-location.

FRANC3D calculates stress intensity factors (K_I , K_{II} , and K_{III}) using the crack opening displacement technique. Values for K_I , crack opening mode stress intensity factors, are calculated at discrete points along the crack front based on the displacement of points on the crack faces some measured distance from the crack front. Referring to Figure 6, K_I is calculated based on the formula¹⁵:

$$K_I = \frac{G}{\kappa + 1} \sqrt{\frac{2\pi}{L}} [4(v_B' - v_D') - (v_C' - v_E')] \quad (1)$$

where G is the shear modulus, κ is $(3-\nu)/(1+\nu)$ in cases of plane stress or $(3-4\nu)$ for plane strain and axisymmetry, ν is Poisson's ratio, ν' is the displacement normal to the crack axis at the indexed point, and L is the distance between point A and points C and E in Figure 6.

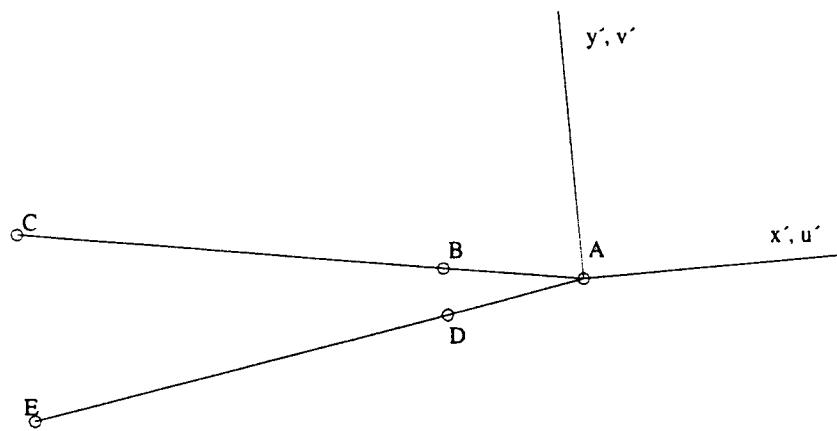


Figure 6: Schematic for crack opening displacement correlation.

Analytical Models

Typical fuselage lap joints on C/KC-135 aircraft consist of two overlapping aluminum (2024-T3) skins of equal thickness (typically 0.063 in.). These are held together by countersunk rivets and are attached to a structural stringer as shown in Figure 7. The faying surfaces on repaired aircraft may also be coated with a sealant intended to prevent or delay corrosion between the layers. The rivets have a typical shank diameter of 0.1875 in. and 100° countersunk heads and are installed on one-inch centers both horizontally and vertically. Evidence from operational aircraft has shown that the upper row of rivets is the critical location in crack initiation. However, in the presence of corrosion pillowing, observed cracks are rarely oriented normal to the hoop stresses caused by fuselage pressurization. Instead, they tend to start in the lower half of the rivet holes and propagate into the pillowed regions between adjacent holes as shown in Figure 4.

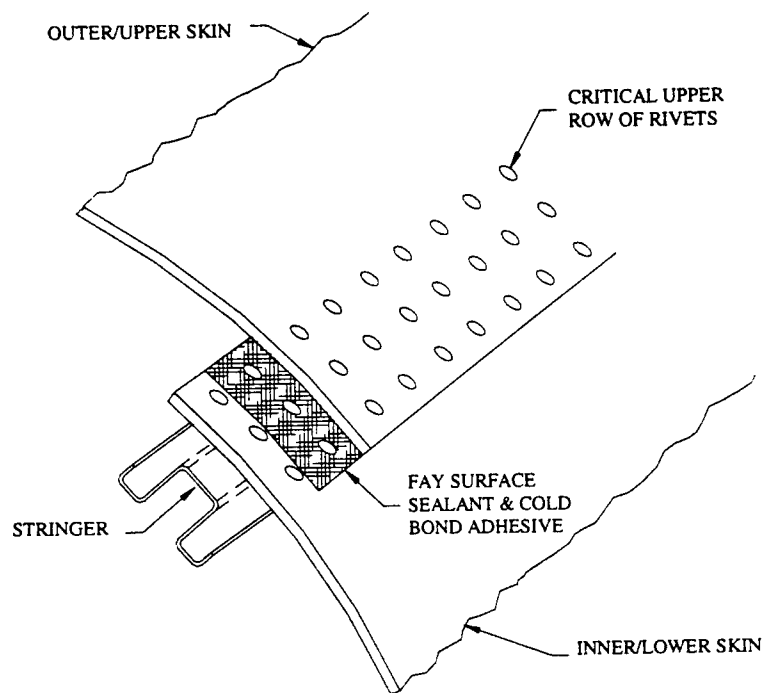


Figure 7: Schematic of C/KC-135 fuselage lap joint.⁵

Preliminary Model

Since the critical cracking location is on the upper skin, it alone was modeled using boundary element methods (BEM) in *FRANC3D*. Because the sealant and adhesive layers must deteriorate for corrosion to occur, neither was included in this analysis. Fuselage curvature effects were assumed negligible and were, therefore, not incorporated into any of the models. The initial model (Figure 8) was created to determine the ability of the *FRANC3D* software to duplicate the general overall stress states achieved by Bellinger, et al.,⁴ under similar loading conditions. This model consisted of two rows and three columns of rivet holes simulating two pillowed areas and their immediate surroundings. Hoop stresses were modeled by a 12 ksi tensile traction along the left surface. The straight shanks of the rivet holes were constrained to have no displacements normal to their surfaces, simulating the presence of rivets. In addition, global z-axis translations were constrained in the countersunk regions of the holes to simulate the

clamping effects of rivet head loads. A uniform pressure was applied to the faying surface to achieve out-of-plane pillowing displacements, and symmetry conditions were imposed on the top and bottom edge surfaces.

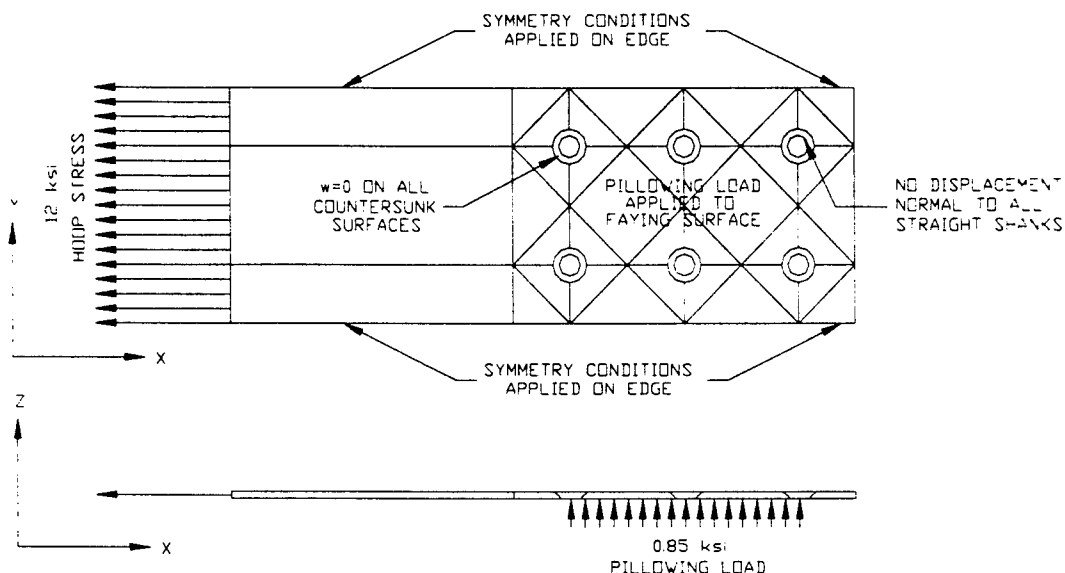


Figure 8: Initial model and boundary conditions used for a qualitative evaluation of *FRANC3D*.

The out-of-plane displacement contours calculated by *FRANC3D* (Figure 9) agreed qualitatively with those obtained by Komorowski, et al., in their elastic-plastic finite element analysis.⁵ Displacements and stresses could not be directly compared due to the limitations imposed by the linear elastic formulation of *FRANC3D*. Unfortunately, the accompanying scale by which to interpret the contour mapping is not included in *FRANC3D* output files. The light areas surrounding the rivet holes correspond to the regions of least displacement, which goes to zero on the countersinks. The darkest areas in the center of the pillowed regions correspond to the regions of maximum z-displacement ($w=0.0048$ inch).

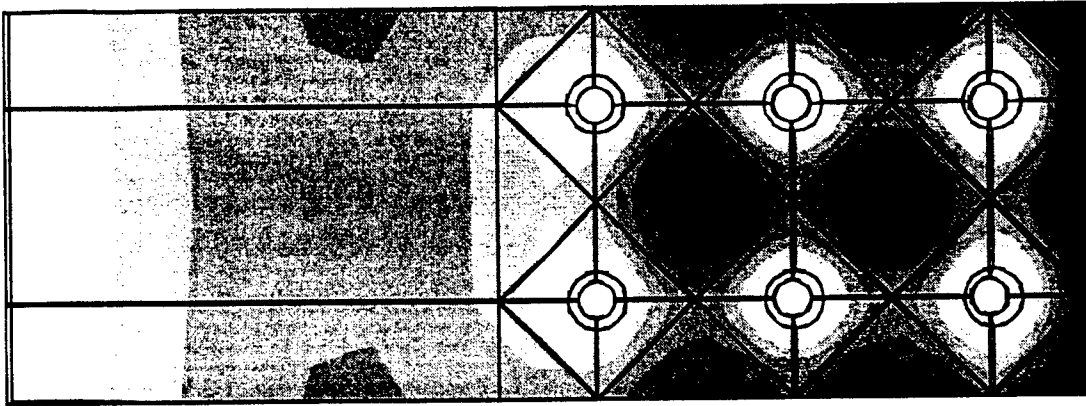


Figure 9: Out-of-plane displacement (w) contours in initial model.

Model for Crack Analysis

Once the analytical capabilities of *BES* were qualitatively confirmed, a second, smaller model was constructed for the fracture mechanics analyses (Figure 10). The second model was used to reduce computational time while duplicating the stress state established by the first model. This model consisted of only one pillowed area representing the region between the upper and middle rows of rivets. Boundary conditions on the rivet hole shanks and countersunk regions were identical to those used in the larger model. Similarly, a 12 ksi traction was applied along the left surface. Symmetry conditions were again applied to the top and bottom edge surfaces with an additional symmetry condition imposed on the right surface. Since *FRANC3D* is only capable of linear elastic analysis, stresses in the model had to be limited to the yield strength of 2024-T3, which is approximately 48 ksi.¹⁸ Pressure was applied to the faying surface such that this criterion was not exceeded. By trial and error, it was found that a uniform pressure of 0.85 ksi on the faying surface, when combined with the other boundary conditions, induced a maximum principal stress of 48 ksi in the uncracked model (von Mises stresses are not reported by *FRANC3D*). The resulting deflection at the center of the pillowed region was 0.0048 inch.

Based on the relationship between central deflection and percent thickness lost to corrosion formulated by Bellinger,⁵ et al., this deflection corresponds approximately to a 2.5% thickness reduction. It is clear from this result that a very small amount of corrosion is sufficient to stress the lap joint material beyond its yield point.

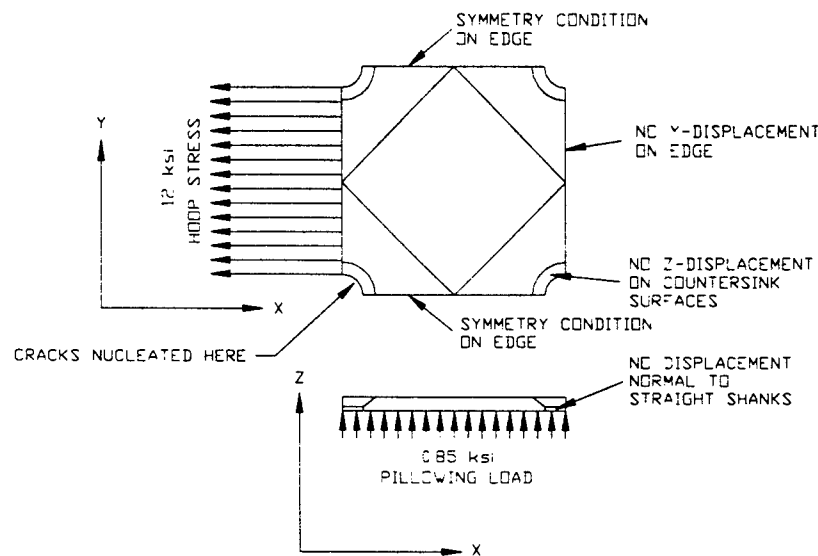


Figure 10: Solid model with boundary conditions used for crack analysis.

The z-axis displacements (w) resulting from the loading and boundary conditions described in Figure 10 are shown in Figure 11. Again, the lightest regions surrounding the rivet holes correspond to zero out-of-plane displacement, while the center of the pillowed region was displaced 0.0048 inch. Under the same conditions, the overall stress field in the model is illustrated in Figure 12. As indicated in the overall picture and the close-up of the stress field surrounding the lower left rivet hole (Figure 13), the majority of the structure is in a state of tensile stress. However, a region of compressive stress, induced by the pillowing deformation, exists in the material near the top of the countersink surface.

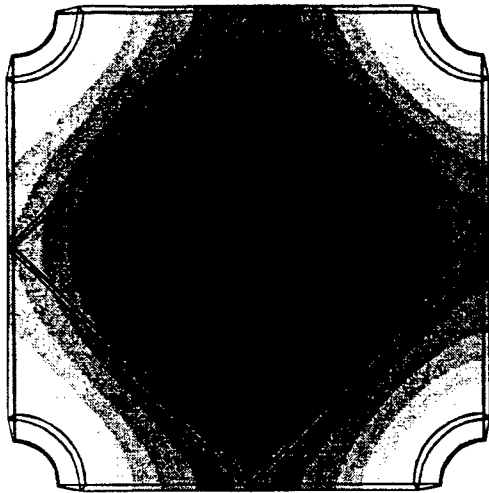


Figure 11: Out-of-plane displacement (w) contours on top surface of crack analysis model.

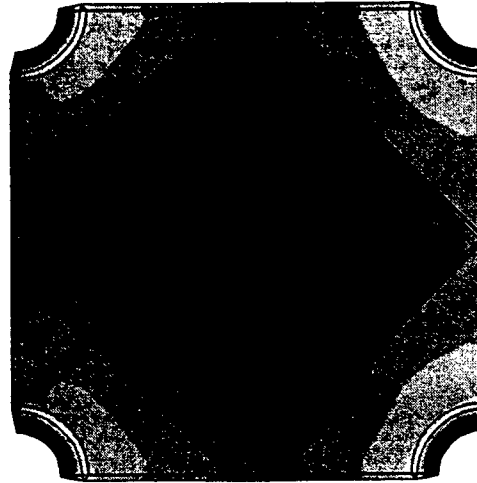


Figure 12: Principal stress field (σ_I) on top surface of crack analysis model.

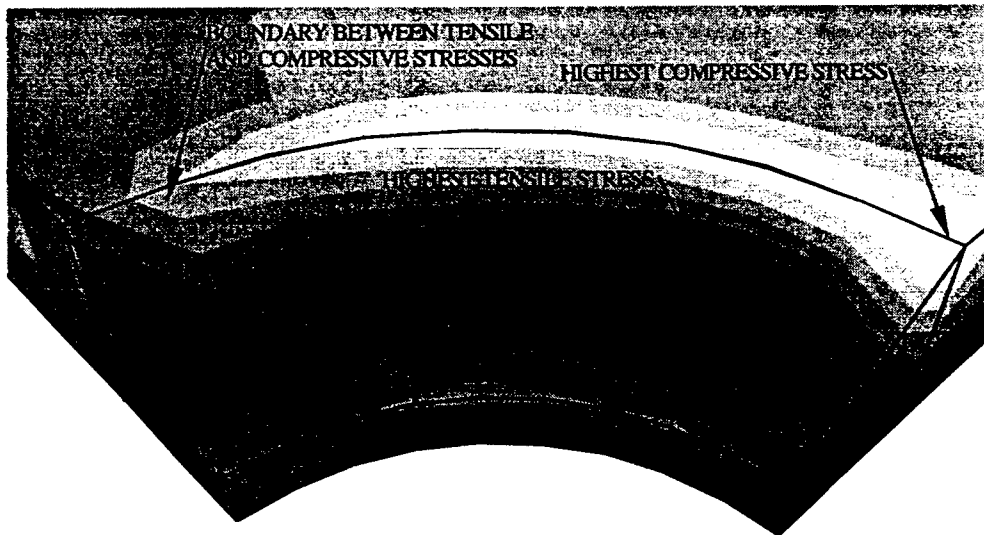


Figure 13: Principal stress field (σ_I) on the surface of the lower left rivet hole in Figure 12.

Analytical Crack Characteristics

All of the cracks analyzed in the countersunk region of the model were planar, circular-front, corner cracks radiating from the rivet hole at the faying surface as shown in Figure 14. Crack front shapes were based on the findings of Willard¹⁹ illustrated in Figure 15 for cracks

near countersunk holes in specimens subjected to uniaxial tension. These afforded the closest available approximation since little research has been done on crack front shapes under the conditions specific to this study. Crack lengths, a , were measured by the distance from the corner of the hole to any point on the round crack face and varied between 0.01 inch and 0.08 inch inclusive. These were non-dimensionalized with respect to the material thickness, $\alpha = a/t$, to give a range of $\alpha = 0.159-1.27$ ($t = 0.063$ inch). Cracks were oriented every 10 degrees between the model edges, resulting in eight angular orientations between 10° and 80° inclusive, as shown in Figure 16. Since the y-axis is normal to the hoop stress, it was used as the reference for crack orientation. Thus, a 10° crack is oriented 10° from the normal to the hoop stress.

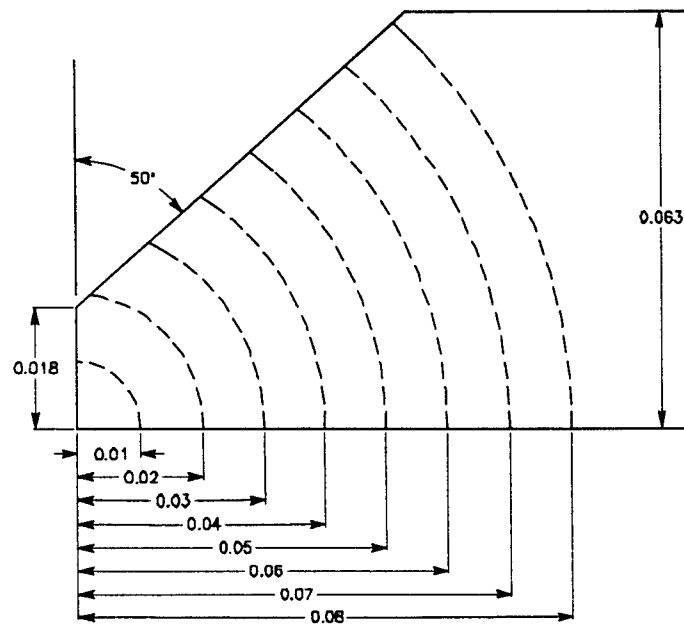


Figure 14: Crack lengths, a , measured from the corner along the faying surface.



Figure 15: Micrograph montage showing corner crack growth near a countersunk hole. Marker bands showing crack front locations at constant intervals are highlighted and numbered.¹⁹

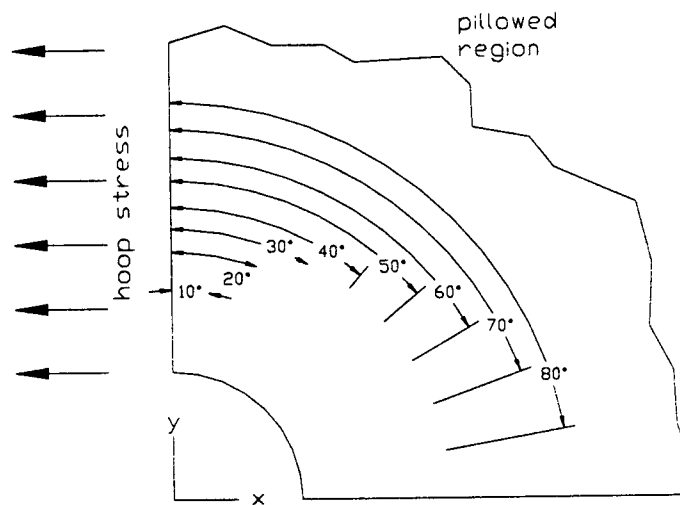


Figure 16: Schematic of crack orientations, β .

Although boundary element methods are relatively insensitive to element shape and aspect ratio on most surfaces, both factors were taken into account on the crack faces near the crack front. The creators of *FRANC3D* recommend that the elements along the crack front be nearly square to assure accurate stress recovery in the stress intensity factor calculations. This was easily achieved through bilinear mapping on the crack faces near the front. A convergence

study was conducted to determine the minimum number of elements required along the crack front. A crack of length $a = 0.01$ inch oriented at $\beta = 10^\circ$ was modeled using 15, 20, and 25 elements along the crack front. The resulting values of K_I are shown in Figure 17. The stress intensity factors calculated from 20 and 25 crack front elements were nearly identical, while those calculated from 15 elements showed minor irregularities. In order to obtain adequate resolution with as few elements as possible, all subsequent cracks were modeled with 20 nearly square elements along their crack fronts as shown in Figure 18.

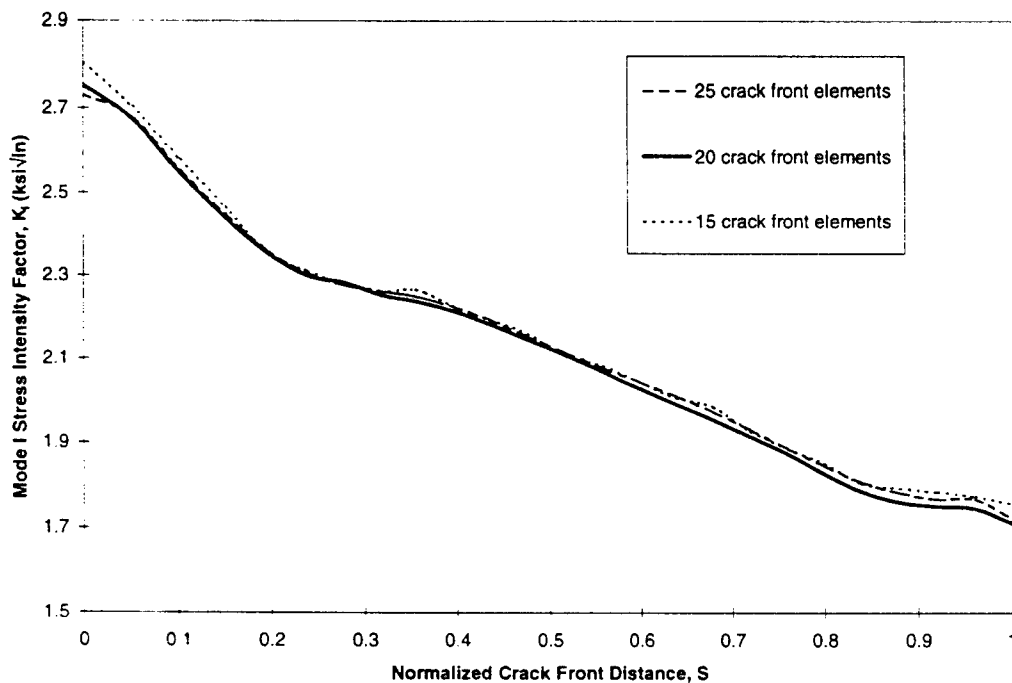


Figure 17: Comparison of K_I distributions calculated with 15, 20, and 25 crack front elements.

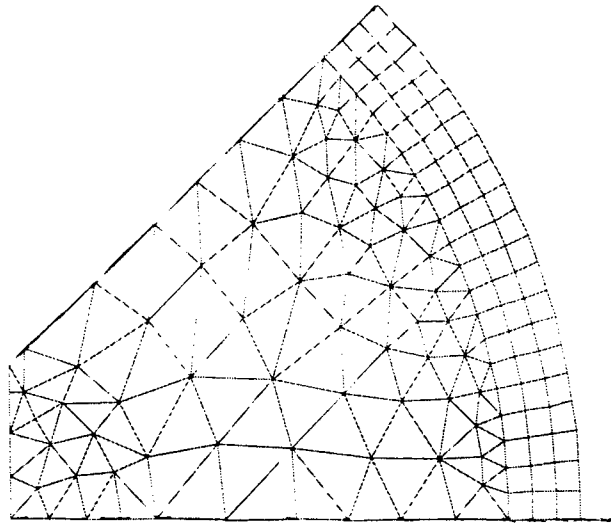


Figure 18: Typical crack face mesh with nearly square elements along the crack front.

In addition to the survey of stress intensity factors for uniform planar cracks in the countersink, cracks in the pillowed region were also considered. Planar through cracks were nucleated from the rivet hole at orientations of 10°, 30°, 60°, and 80° as defined above, and extended 0.05 inch beyond the countersink. Subsequent cracks were then formed using two different methods. The first was uniform planar extension, resulting in a change of length only. This allowed for an analysis similar to that done in the countersink. The second method was through the use of an automated calculation in *FRANC3D* based on a logarithmic formulation similar to Paris's Law.¹⁶ The equation used by *FRANC3D* to calculate the extension was:

$$\delta = \delta_{\max} \times \frac{K_I^b}{K_{I\max}^b} \quad (2)$$

where δ_{\max} is the user-specified maximum extension of the crack tip and b is a parameter obtained from the relationship between stress intensity factors and the growth of through cracks in two-dimensional structures under plane strain. The terms K_I and $K_{I\max}$ are the stress intensity factor at a specific point and the maximum stress intensity factor on the crack front, respectively.

The default value of $b = 1$ was left unchanged because insufficient information is available as to the growth rates of these cracks to make a more accurate determination. The method used to determine the *direction* of crack propagation must be chosen by the user from the following theories: the maximum tangential stress theory, the maximum strain energy release rate, and the minimum strain energy density.¹⁶ Again, due to limited information about these cracks, the default selection of the maximum tangential stress theory was used.

Results

Stress intensity factors were calculated for cracks throughout the countersunk region surrounding the rivet hole. These were grouped for analysis by both angular orientation and crack length. For each angular orientation, the stress intensity factors for all cracks analyzed at that orientation were plotted versus the normalized crack front distance (Figures 19 through 26). The normalized crack front distances are arc lengths along the crack front measured from the faying surface and normalized with respect to the total crack front length. These plots were used to visualize the way in which the stress intensity varied with crack length. Similarly, three-dimensional plots were created to compare how stress intensities varied with angular orientation for cracks of a specific length (Figures 27 through 34).

Analysis of the variation of K_I with crack length revealed several interesting trends. First, it was observed that in all cases, the stress intensity factor at the faying surface, corresponding to a normalized position of 0.0, increased with crack length. This was to be expected from basic fracture mechanics since stress intensity factors are proportional to crack length in a given stress state. Also of interest is the manner in which the stress intensities decrease rapidly in the upper portions of the crack front, i.e. near the countersink surface, for many of the orientations. This

does not occur in the 10° cracks, which are most nearly normal to the direction of the hoop stress, as the curve maintains a characteristic shape for all crack lengths. At this orientation ($\beta = 10^\circ$), the tensile hoop stress is sufficient to overcome the compressive loads created by the pillowing displacement (Figure 19). However, as the crack orientations become more nearly parallel to the hoop stresses, the drop in stress intensities near the upper surface of the countersink becomes increasingly pronounced. This is consistent with the compressive stress field near the upper surface surrounding the rivet hole as shown in Figure 13. For cracks oriented at 30° or greater, the stress intensities near the top of the crack front go to zero for the larger cracks. In fact, the numerical analysis of these cracks resulted in negative values for stress intensity factors, which are physically impossible. This was simply the result of the analytical crack faces being compressed through each other resulting in negative crack opening displacements. Under conditions of compressive stress that would cause these negative displacements, K_I has no meaning. Therefore, any negative stress intensity values were set to zero. Although Figure 13 reveals only the surface stresses, the extent to which the compressive stress field extends below the surface is illustrated by the crack front stress intensity factors being zero.

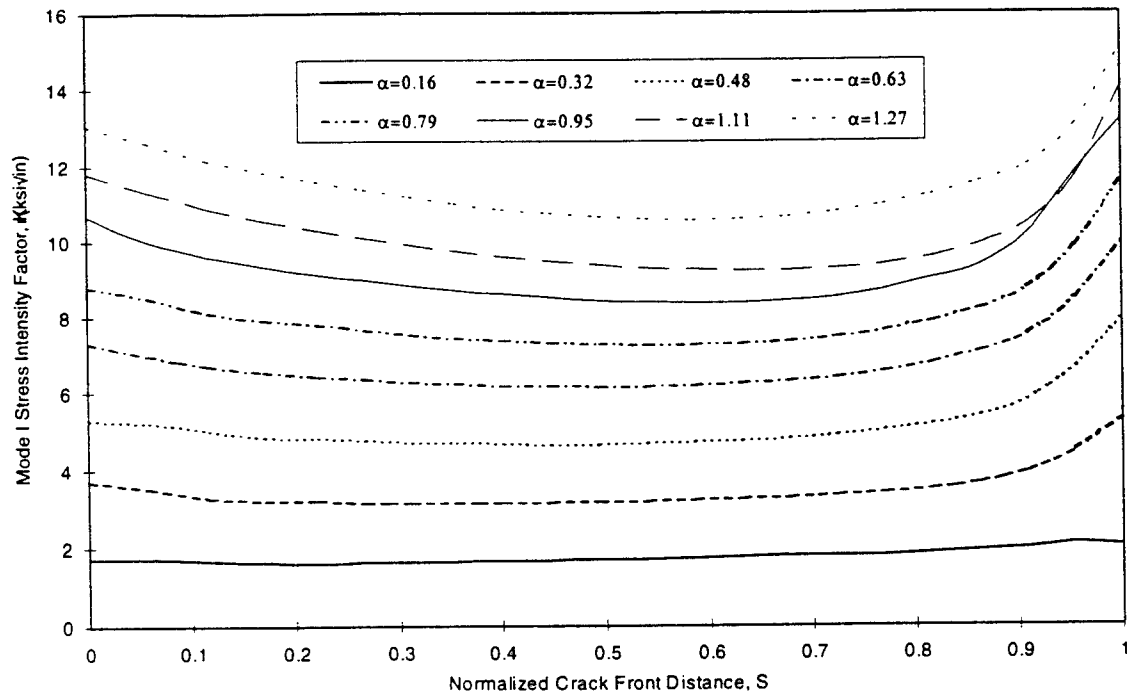


Figure 19: Stress intensity factors vs. normalized crack front distance for cracks oriented at 10°.

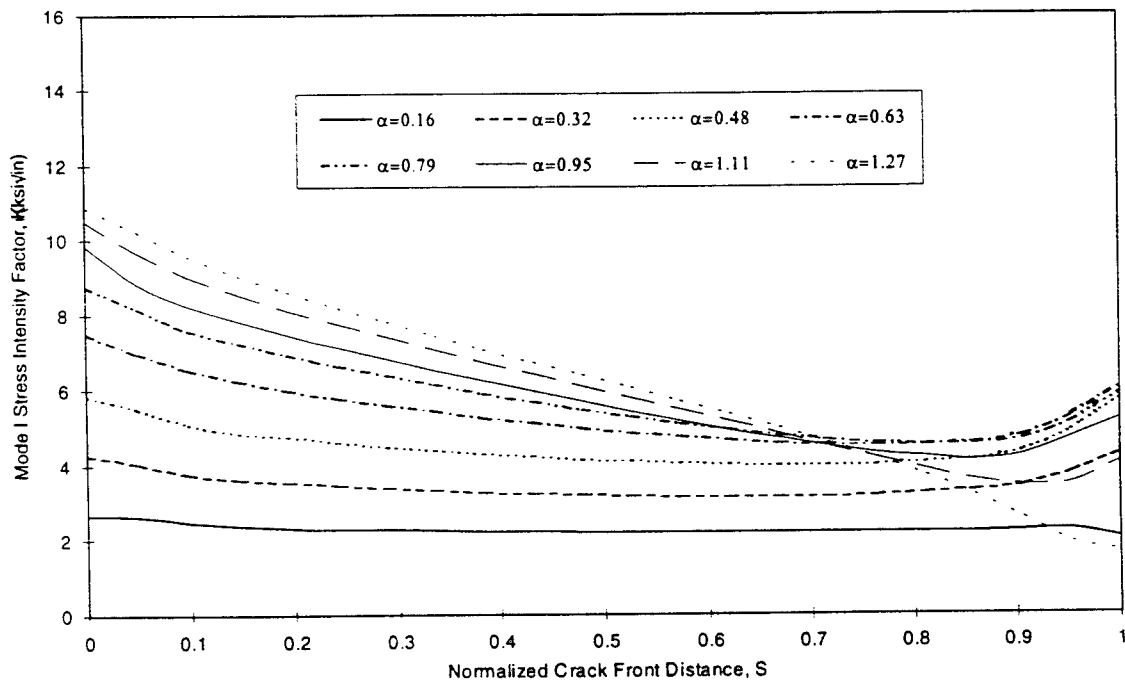


Figure 20: Stress intensity factors vs. normalized crack front distance for cracks oriented at 20°.

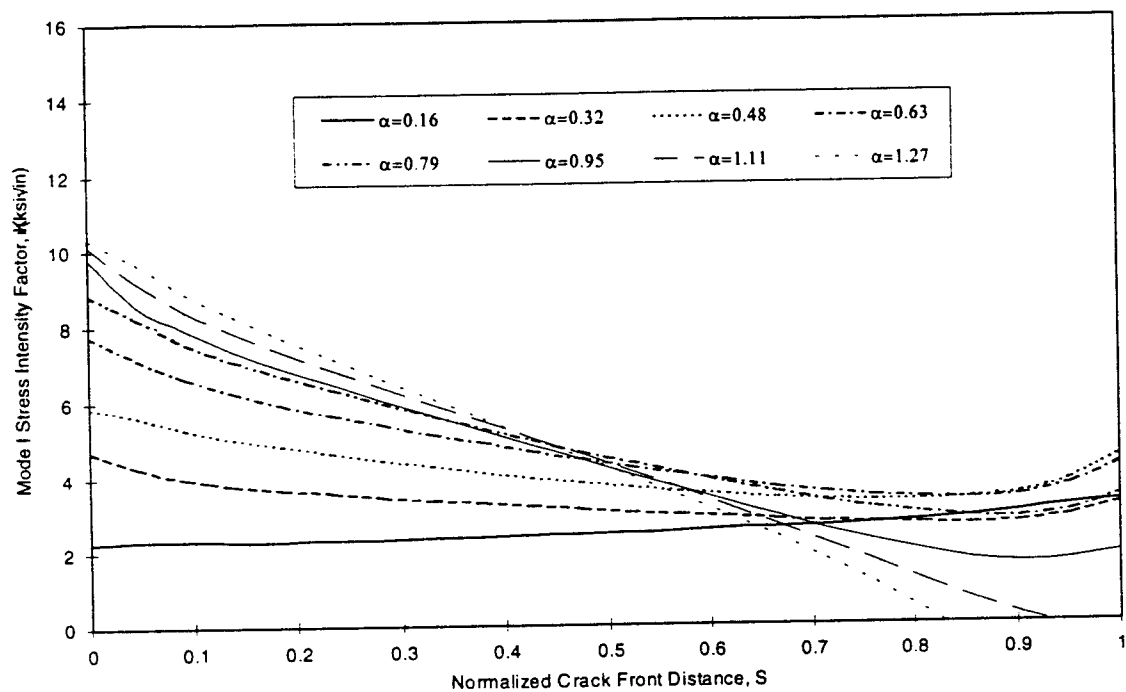


Figure 21: Stress intensity factors vs. normalized crack front distance for cracks oriented at 30° .

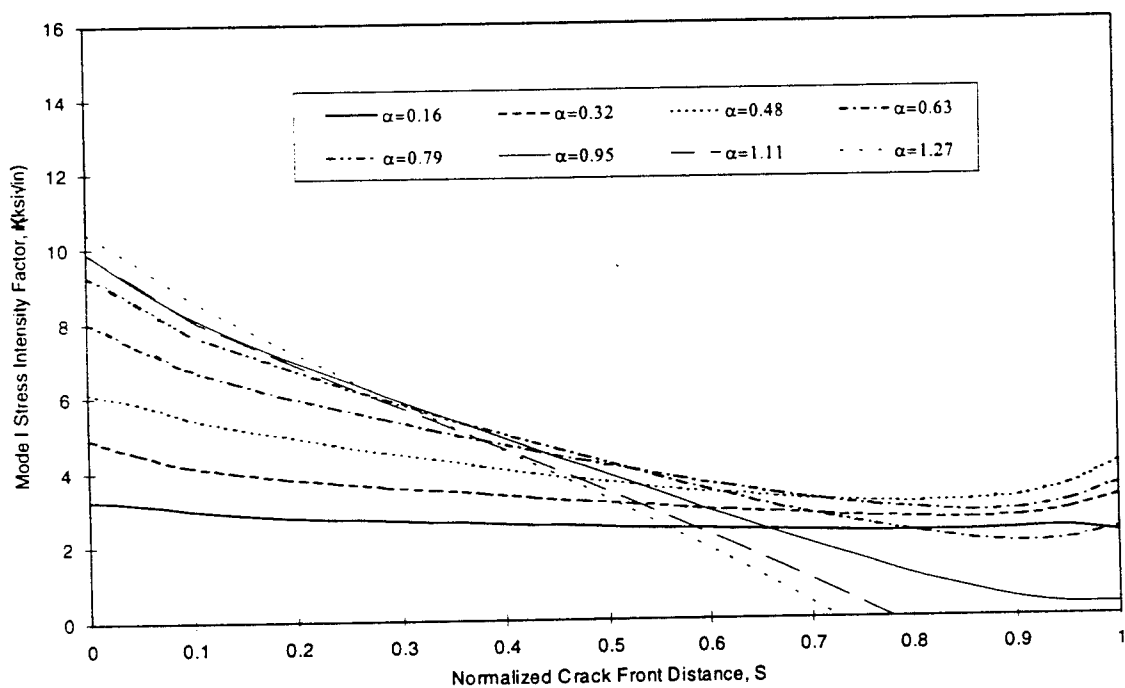


Figure 22: Stress intensity factors vs. normalized crack front distance for cracks oriented at 40° .

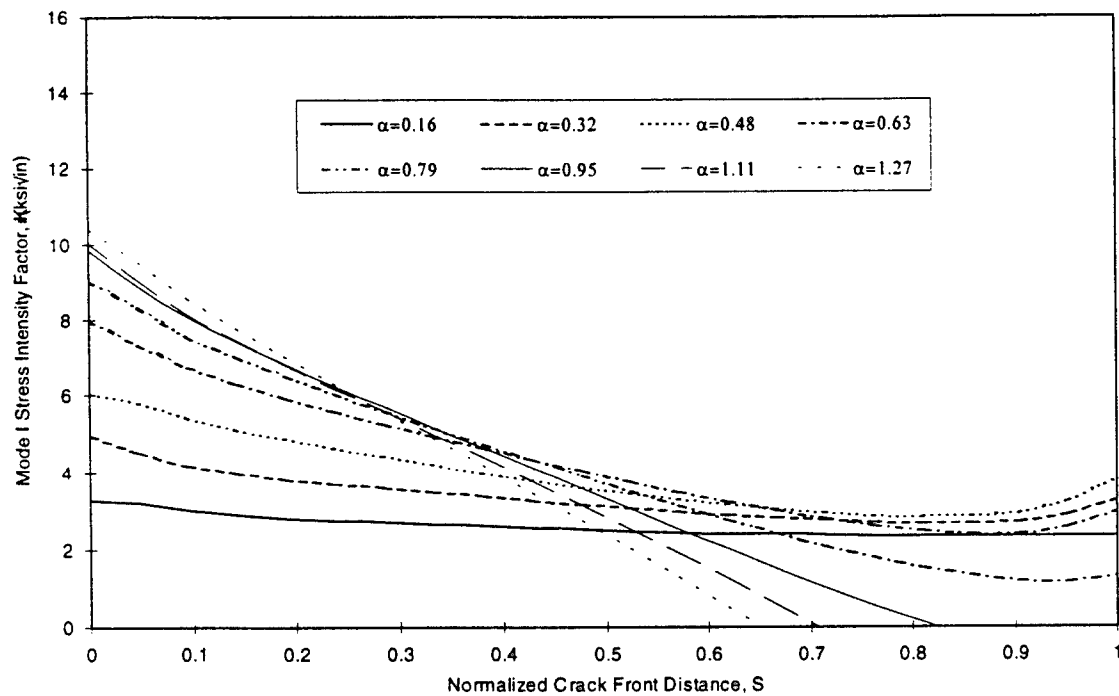


Figure 23: Stress intensity factors vs. normalized crack front distance for cracks oriented at 50° .

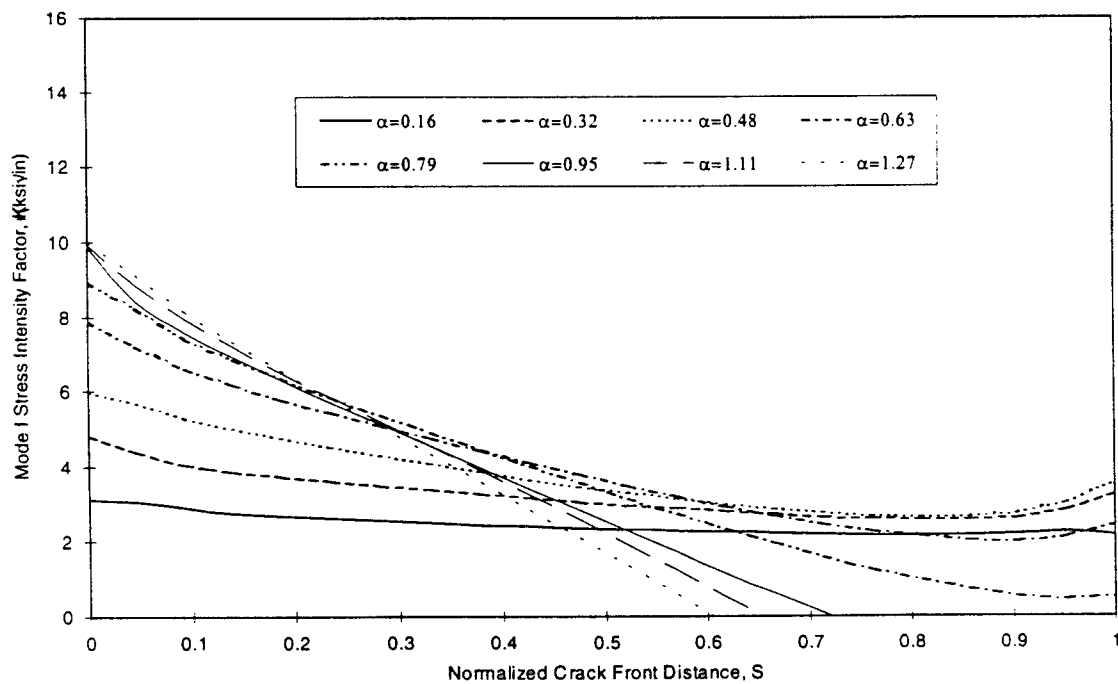


Figure 24: Stress intensity factors vs. normalized crack front distance for cracks oriented at 60° .

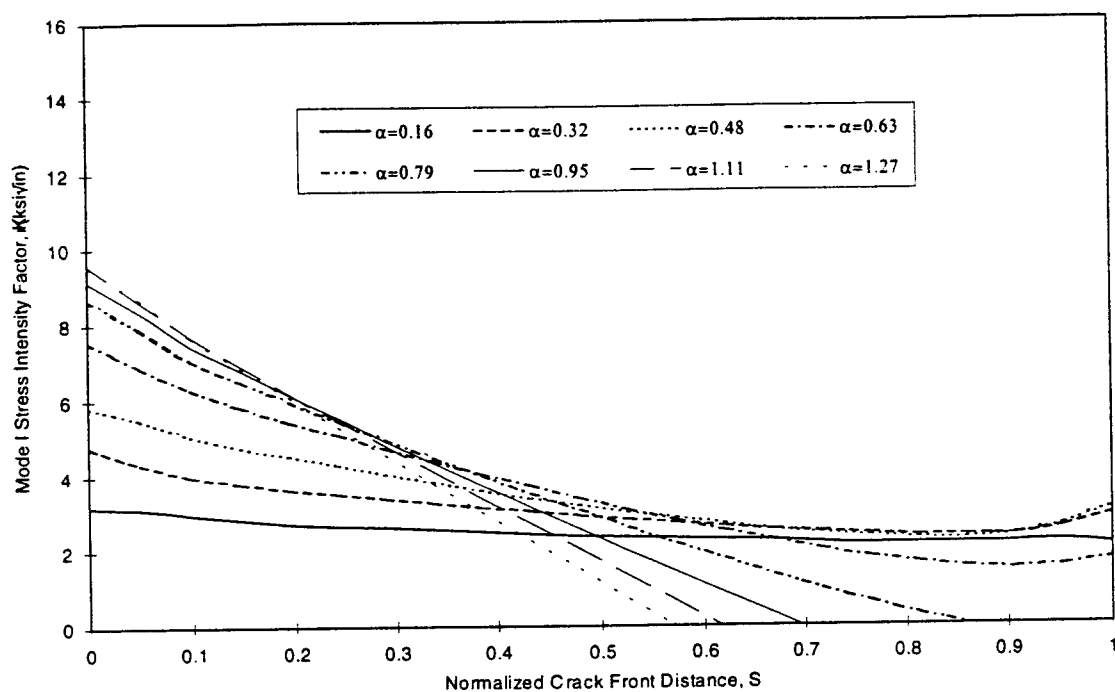


Figure 25: Stress intensity factors vs. normalized crack front distance for cracks oriented at 70° .

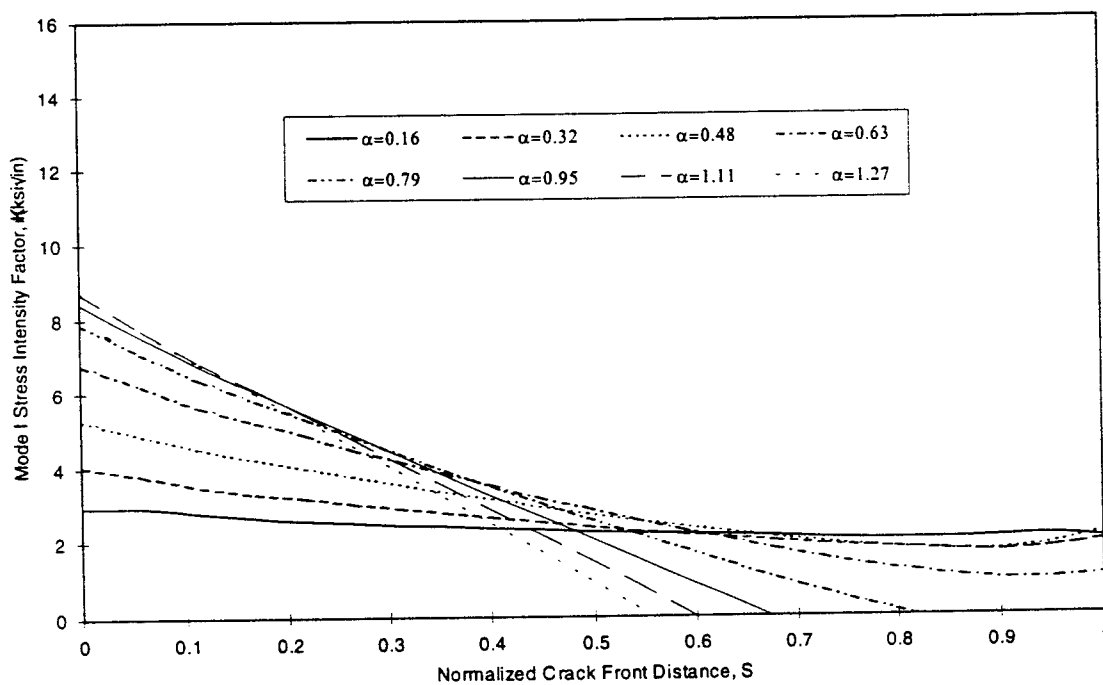


Figure 26: Stress intensity factors vs. normalized crack front distance for cracks oriented at 80° .

Examination of the three-dimensional plots of K_I versus crack front location and crack orientation revealed that the stress intensity factors were relatively uniform in smaller cracks ($\alpha = 0.16$) over the range of angles (Figure 27). These were corner cracks confined to the straight shank portion of the rivet hole, not yet having broken through the countersunk surface. The first cracks that intersected the countersink ($\alpha = 0.32$) began to display larger K_I variations with respect to orientation. This can first be seen in Figure 28. Stress intensity factors were nearly uniform across the faying surface for all angles at $\alpha = 0.32$. The 10° crack of this length maintained the same K_I distribution as seen in Figure 27, although magnified due to the increased crack length. The qualitatively similar K_I distributions for the 10° crack were common throughout the range of crack lengths, with the overall level of stress intensity increasing as α increased. Other crack orientations, particularly in the range from 30° to 80° , displayed a marked decrease in K_I toward the countersunk surface. Beginning with cracks of length $\alpha = 0.79$, those with orientation angles of more than 60° developed regions of zero stress intensities near the countersink. As the crack length is increased, the region where K_I is equal to zero was encountered by cracks of just over 20° (Figure 34).

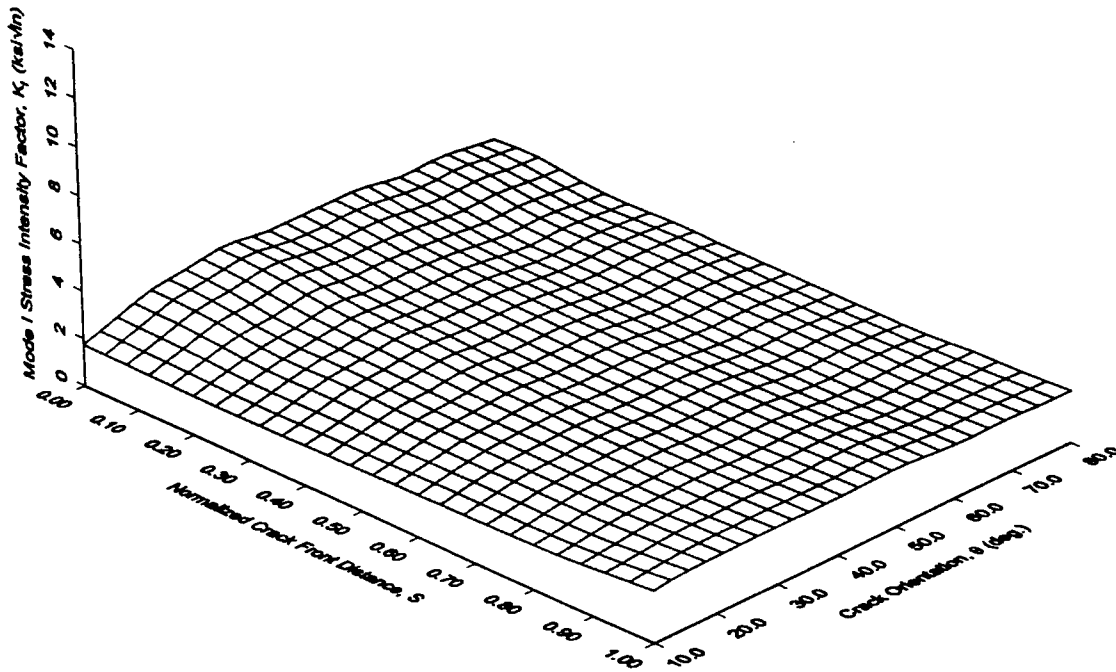


Figure 27: Stress intensity factors vs. normalized crack front distance for all cracks $\alpha=0.16$.

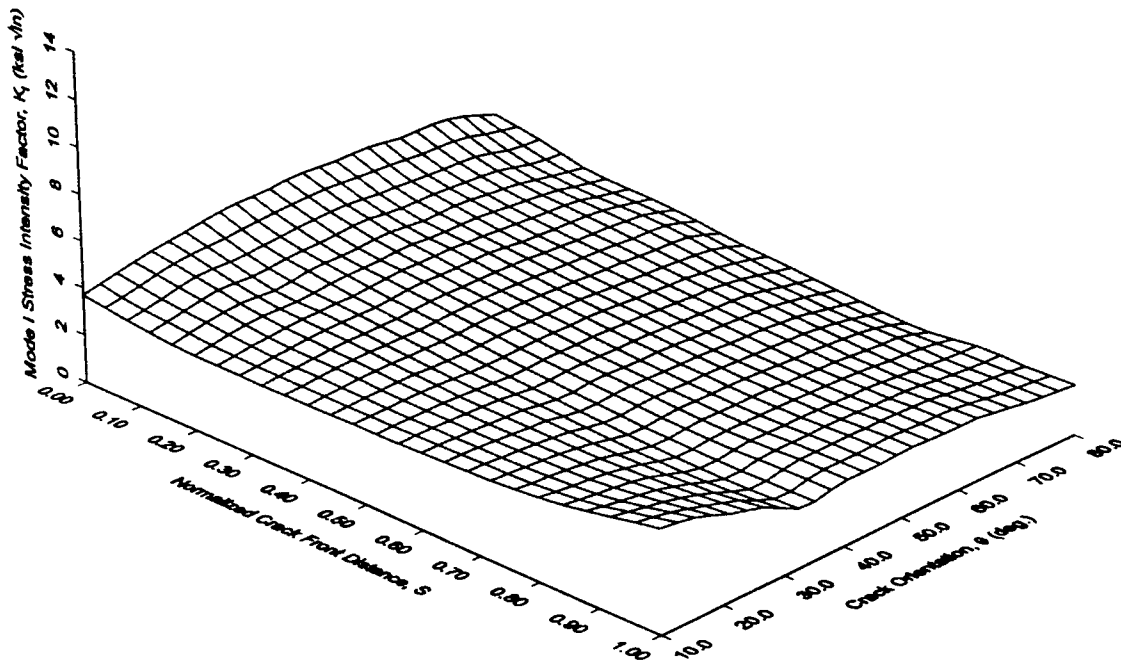


Figure 28: Stress intensity factors vs. normalized crack front distance for all cracks $\alpha=0.32$.

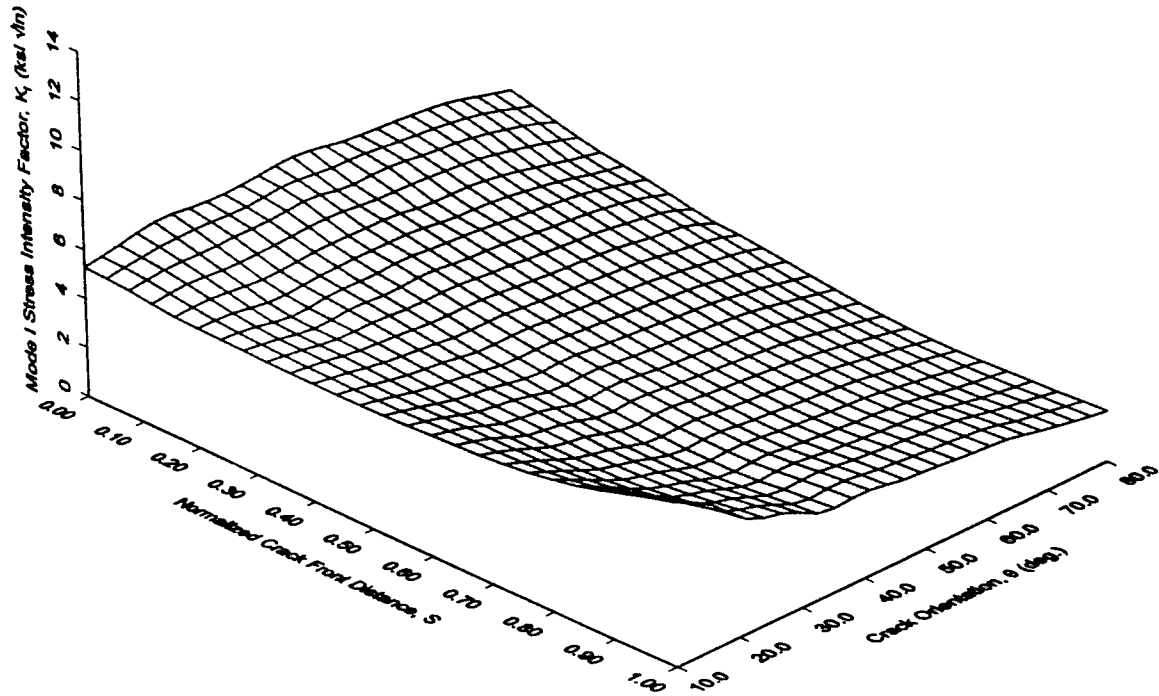


Figure 29: Stress intensity factors vs. normalized crack front distance for all cracks $\alpha=0.48$.

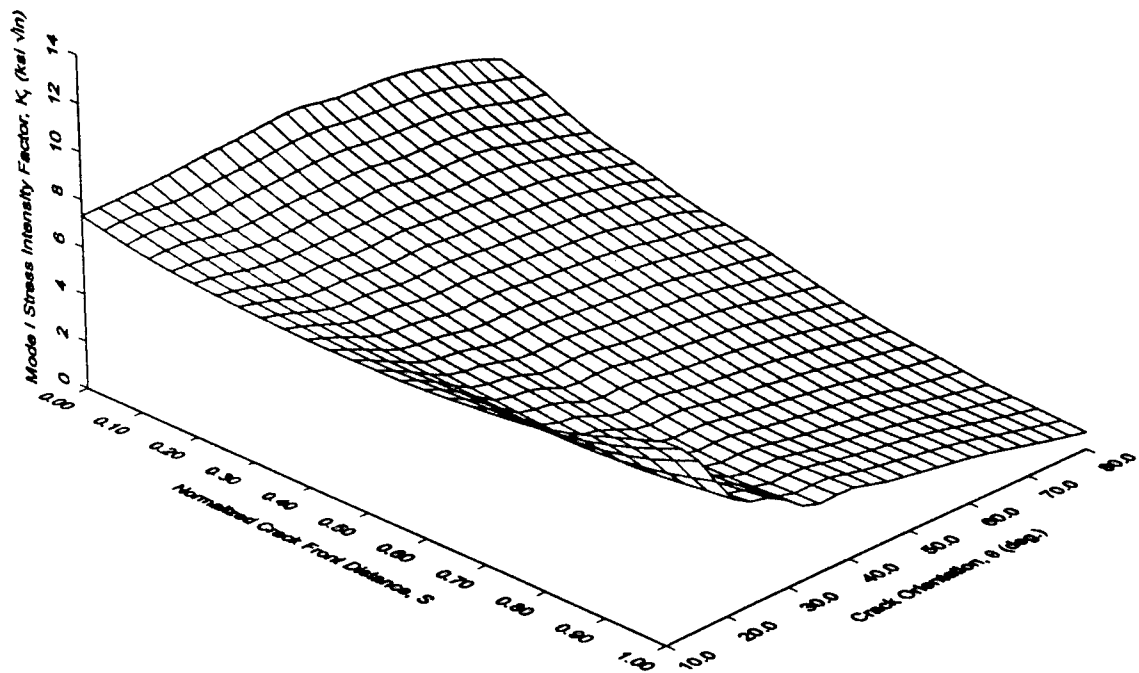


Figure 30: Stress intensity factors vs. normalized crack front distance for all cracks $\alpha=0.63$.

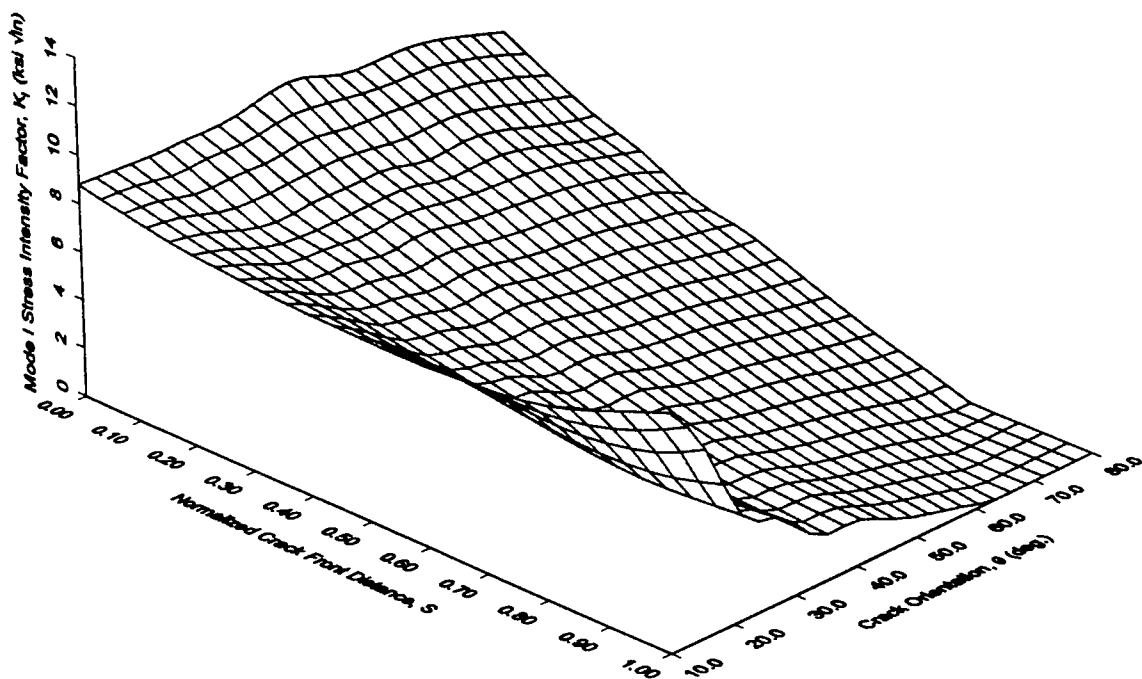


Figure 31: Stress intensity factors vs. normalized crack front distance for all cracks $\alpha=0.79$.

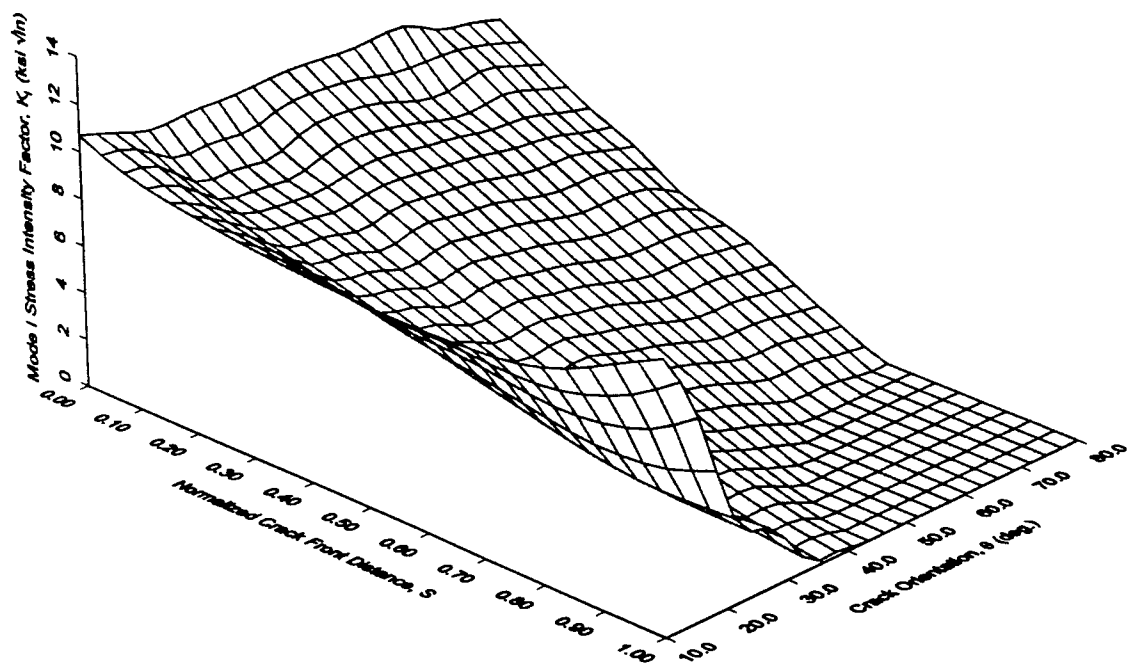


Figure 32: Stress intensity factors vs. normalized crack front distance for all cracks $\alpha=0.95$.

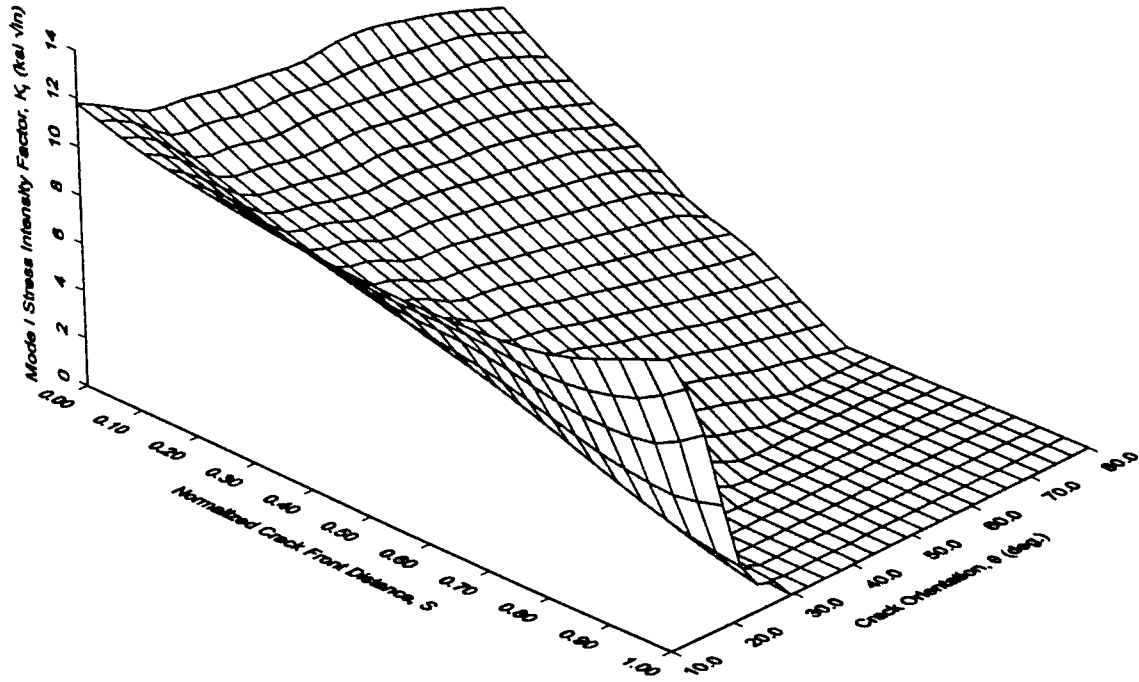


Figure 33: Stress intensity factors vs. normalized crack front distance for all cracks $\alpha=1.11$.

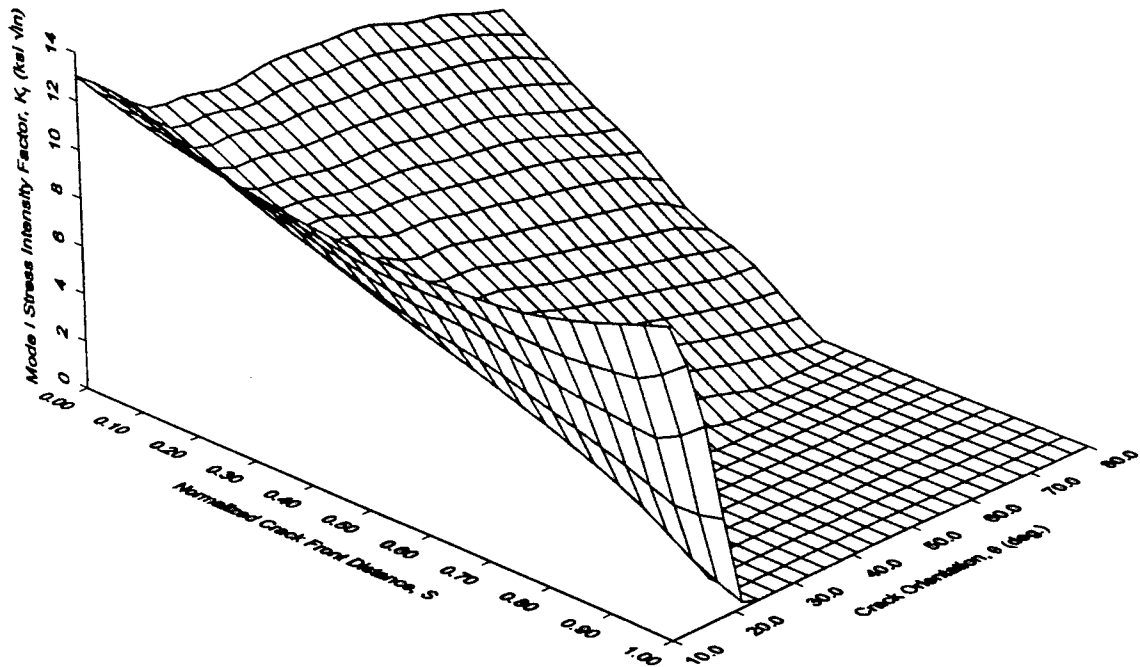


Figure 34: Stress intensity factors vs. normalized crack front distance for all cracks $\alpha=1.27$.

A comparison of stress intensity factors at the faying surface for the smallest cracks analyzed ($\alpha = 0.16$), reveals that the 10° crack had the lowest value of K_I . This suggests that if material imperfections were present all along the rivet hole edge, those oriented away from the normal to the hoop stress would be more likely to nucleate cracks. This condition would explain the absence of cracks radiating toward the next adjacent rivet hole, and instead propagating into the pillowed region. Corrosion pillowing alters the stress state to such an extent that the hoop stress is no longer the driving force in crack initiation. However, it should be noted that the stress intensity factors for cracks at 10° exceed those for cracks at any other orientation for crack lengths beyond $\alpha = 0.952$. This indicates that pre-existing cracks of this length or greater at an orientation of 10° would tend to propagate faster than those at other orientations. It is important to remember that this study was limited to a linear elastic analysis. This dictated the maximum allowable stresses, and consequently the maximum pillowing deflection. An elastic-plastic analysis should be conducted to determine the effects of increased corrosion on crack nucleation sites.

The data also offer some explanation for the high aspect ratio cracks on lap joint faying surfaces. High stress intensity factors at the faying surface indicate that cracks will propagate readily in that region. However, as the stress intensities tend to decrease and eventually go to zero in the compressive stress field near the countersink surface, crack growth would be arrested before becoming through cracks. The locus of points where stress intensity factors go to zero defines a neutral surface separating tensile stress near the faying surface and compressive stress near the countersink or outer surface. The maximum height to which a faying surface crack could grow would be limited to some level below this neutral surface, since even positive stress intensities are not sufficient for crack propagation. At some point below the neutral surface,

stress intensities could reach the minimum or critical stress intensity factor at which crack growth would occur. Again, because of the elastic limitations of *FRANC3D*, the critical stress intensity factor for 2024-T3 was never attained. This was because the maximum principal stress was located outside the region of crack investigation. An elastic-plastic analysis would allow for higher levels of pillowing that could cause the stress intensity at some point to reach K_{Ic} .

Attempts to model the propagation of planar through cracks beyond the countersink were unsuccessful. Portions of the crack front remained in regions of compressive stress beyond the countersink diameter, which can be seen in Figure 13. This resulted in the calculation of negative values for K_I along the upper segment of the crack fronts. Since negative stress intensities are physically impossible and would result in the extension of only the lower segments of the cracks, *FRANC3D* was unable to propagate these cracks. Modeling of through crack propagation was attempted because it was the natural extension of the round corner cracks modeled in the rivet holes. However, given the findings of this part of the study and the fact that high aspect ratio surface cracks are more commonly found in actual pillowed lap joints, surface cracks initiated on the faying surface would be the subject of future studies.

Concluding Remarks

It was the aim of this study to determine the characteristics of cracks in countersunk rivet holes under the conditions imposed by corrosion-induced pillowing in aircraft fuselage lap joints. It was hoped that this investigation would lead to a better understanding of cracks observed in operational aircraft. These include cracks that nucleate at orientations other than normal to the hoop stress (Figure 4) and high aspect ratio cracks along the faying surface (Figure 3). The

evidence collected tends to offer explanations to both of these phenomena, and can be summarized as follows:

1. Stress intensity factors for small cracks ($\alpha = 0.16$) are greater at orientations other than 10° , offering more favorable crack nucleation sites than those oriented normal to the hoop stress.
2. A compressive stress field exists within the material near the countersink surface and extends beyond the diameter of the countersink into the pillowed region. This prevents corner cracks from breaking through the upper surface and promotes the high aspect ratio cracks found in actual pillowed lap joints.

An elastic-plastic analysis should be done to investigate the effects of higher levels of corrosion and the resulting increase in pillowing deformation. Increased pillowing would almost certainly create greater tensile stresses in the material surrounding the rivet hole. The resulting increase in stress intensity may give a clearer indication of why cracks tend to initiate where they do. An elastic-plastic analysis would also allow the stress intensity factors to reach the critical value over some part of the crack front. From this, the region susceptible to crack growth could be determined and the upper limit of faying surface crack growth could be defined.

For crack propagation in the pillowed region to be successfully modeled, surface cracks should be nucleated along the faying surface at various locations, not necessarily at the corner of the rivet hole. These cracks could then be propagated into the pillowed region or toward the rivet hole as dictated by the stress intensities along the crack fronts. This method of crack propagation could be done with a *FRANC3D* elastic analysis or in subsequent elastic-plastic analyses.

The stress state in the material surrounding the rivet hole suggests that crack propagation in this region is a mixed mode phenomenon. Future studies should also include the analysis of Mode II and possibly Mode III stress intensities.

Acknowledgements

This report is based in part on a Master of Science thesis *Fracture Mechanics Analysis of Corroded Aircraft Fuselages Lap Joints* presented by Douglas Welch at the University of Oklahoma. Mr. Clare Paul and Dr. Thomas Mills at Wright Laboratories provided valuable encouragement and suggestions in the course of this study. Dr. Jerzy Komorowski at NRC-Canada supplied the photographs of the pillowed lap joint and its associated cracks. Dr. Robert Piascik and Mr. Scott Willard of the NASA Langley Research Center provided the photograph of the countersink crack fronts. The contributions of all these people has been crucial to the results presented here.

References

1. *Aging of U.S. Air Force Aircraft: Final Report*, Committee on Aging of U.S. Air Force Aircraft, National Materials Advisory Board, National Research Council, National Academy Press, NMAB-488-2, Washington, D.C., 1997, pp. 13-21.
2. Lincoln, J.W., "The USAF Approach to Attaining Structural Integrity of Aging Aircraft," *Structural Integrity in Aging Aircraft*, 1995 ASME International Mechanical Engineering Congress and Exposition, pp. 9-19.
3. Komorowski, J.P., N.C. Bellinger, and R.W. Gould, "The Role of Corrosion Pillowing in NDI and in the Structural Integrity of Fuselage Joints," *Fatigue in New and Aging Aircraft – Proceedings: 19th Symposium of the International Committee on Aeronautical Fatigue*, Edinburgh, 16-20 June 1997. (To be published)
4. Bellinger, N.C., and J.P. Komorowski, "Corrosion Pillowing Stresses in Fuselage Lap Joints," *AIAA Journal*, Vol. 35, No. 2, February 1997, pp. 317-320.
5. Bellinger, N.C., J.P. Komorowski, and S. Krishnakumar, *Numerical Modeling of Pillowing Due to Corrosion in Fuselage Lap Joints*, National Research Council – Canada, Institute for Aerospace Research, LTR-ST-2005, April 1995.

6. Bellinger, N.C., S. Krishnakumar, and J.P. Komorowski, "Modeling of Pillowing Due to Corrosion in Fuselage Lap Joints," *Canadian Aeronautics and Space Journal*, Vol. 40, No. 3, September 1994, pp. 123-130.
7. Smith, F.W. and T.E. Kullgren, *Theoretical and Experimental Analysis of Surface Cracks Emanating from Fastener Holes*, AFFDL-TR-76-104, Air Force Flight Dynamics Laboratory, Wright-Patterson AFB, February 1977.
8. Raju, I.S. and J.C. Newman, Jr., "Stress Intensity Factors for Two Symmetric Corner Cracks," *Fracture Mechanics*, ASTM STP 677, 1979, pp. 411-430.
9. Newman, J.C., Jr. and I.S. Raju, "Stress-Intensity Factor Equations for Cracks in Three-Dimensional Finite Bodies," *Fracture Mechanics: Fourteenth Symposium - Volume I: Theory and Analysis*, ASTM STP 791, 1983, pp. I-238 - I-265.
10. Nishioka, T. and S.N. Atluri, "An Alternating Method for Analysis of Surface-Flawed Aircraft Structural Components," *AIAA Journal*, 21(5), 1983, pp. 749-757.
11. Tan, P.W., I.S. Raju, K.N. Shivakumar and J. C. Newman, Jr., "Evaluation of Finite-Element Models and Stress-Intensity Factors for Surface Cracks Emanating from Stress Concentrations," *Surface-Crack Growth: Models, Experiments, and Structures*, ASTM STP 1060, 1990, pp. 34-48.
12. Shivakumar, K.N. and J.C. Newman, Jr., "Stress-Intensity Factors for Large Aspect Ratio Surface and Corner Cracks at a Semi-Circular Notch in a Tension Specimen," *Engineering Fracture Mechanics*, 38(6), 1991, pp. 467-473.
13. Zhao, W. and X.R. Wu, "Stress Intensity Factors for Corner Cracks at a Semi-Circular Notch Under Stress Gradients," *Engineering Fracture Mechanics*, 13(4), 1990, pp. 347-360.
14. Tan, P.W., C.A. Bigelow, P.E. O'Donoghue and S.N. Atluri, "Stress Intensity Factor Solutions for Cracks at Countersunk Rivet Holes Under Uniaxial Tension," *Durability and Structural Reliability of Airframes - Proceedings of ICAF '93*, Stockholm, Sweden, A.F. Blom, ed., pp. 1191-1215.
15. Lim, I.L., I.W. Johnston, and S.K. Choi, "Comparison Between Various Displacement-Based Stress Intensity Factor Computation Techniques," *International Journal of Fracture*, Vol. 58, 1992, pp. 193-210.
16. Ingraffea, A.R. et al., *FRANC3D: A Three Dimensional Fracture Analysis Code - Concepts and User's Guide*, February 1995.
17. Banerjee, P.K., and R. Butterfield, *Boundary Element Methods in Engineering Science*, London: McGraw-Hill Book Company Ltd., 1981.

18. Military Handbook: Metallic Materials and Elements for Aerospace Vehicle Structures, Volume 1, MIL-HDBK-5G, 1 November 1994.
19. Willard, S.A., "Use of Marker Bands for Determination of Fatigue Crack Growth Rates and Crack Front Shapes in Pre-Corroded Coupons," NASA Contractor Report, Lockheed-Martin Engineering and Sciences Company, Hampton, Virginia. (To be published)

DECONVOLUTION OF
THE SPACE-TIME RADAR SPECTRUM

Armando B. Barreto
Assistant Professor
Department of Electrical and Computer Engineering

Florida International University
Center For Engineering & Applied Sciences - EAS 3942
10555 West Flagler Street
Miami, Florida, 33174

Final Report for:
Summer Research Extension Program
Wright Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Radar Branch
Radio Frequency Division
Avionics Directorate
Wright Laboratory
Wright-Patterson Air Force Base, OH.

December 1997

DECONVOLUTION OF THE SPACE-TIME RADAR SPECTRUM

Armando B. Barreto
Assistant Professor
Department of Electrical and Computer Engineering
Florida International University

Abstract

This report summarizes the continued exploration of the use of deconvolution on the Doppler-Azimuth Spectrum of signals collected by an airborne radar array. Deconvolution is applied to the Azimuth-Doppler Spectrum (ADS) with the objective of emphasizing the information associated with valid targets, while de-emphasizing unwanted signal components such as irrelevant background returns, or clutter, and interference radiated to obscure valid radar returns, or jammer. Alternative forms of implementation for the concept of deconvolution are explored that prove less computationally intensive than the Non-Negative Least Squares (NNLS) method, initially proposed for this purpose in the preceding Summer Faculty Research Program project. Van Cittert's method for iterative constrained deconvolution was found to be a viable candidate method. Both the NNLS and Van Cittert's method have been taken from their prototype implementations in an interpreted simulation environment (Matlab ®), to portable ANSI C implementations. The advantages and disadvantages observed for the methods in a number of spectral scenarios studied are summarized.

DECONVOLUTION OF THE SPACE-TIME RADAR SPECTRUM

Armando B. Barreto

Introduction

Modern airborne radar systems incorporate a number of radar receiver elements to enable them to perceive the angle of incidence of a returned signal. So, in addition to the determination of the range or distance between the instrument and the scatterer and the relative velocity between the two, detected by Doppler effect [5][9][12], the data collected by a multi-element radar system encodes the azimuthal position of the scatterer[4]. This added information can be used to further distinguish between wanted and unwanted components in the signals received by the radar array. For example, a valid moving target will normally be represented by a return of unique Doppler shift (indicative of the relative velocity of the target and the radar instrument), from a unique azimuth, with respect to the broadside of the radar array. In contrast, a broadband or *barrage jammer* signal will be perceived also as arriving from a single azimuth (the azimuth of the jamming source with respect to the broadside of the array), but it will be sensed at many Doppler frequencies. Another unwanted component in the signals received by the airborne array is defined by the irrelevant returns associated with the background illuminated by the radar, or *clutter*. Because there may be returns from background all around the radar instrument, this form of interference may be perceived as if arriving to the instrument from all azimuths. Since the background scatterers are static, the Doppler characteristic of clutter returns varies as function of azimuth, which is determined by the velocity of the radar platform.

Given the potential differentiation between valid target returns, clutter returns and jamming interference, the signal processing challenge is to develop techniques capable of emphasizing the target data and simultaneously eliminate or attenuate the unwanted signal components. Such processing scheme will enhance the efficiency and reliability of radar systems to detect and track moving targets.

Discussion of the problem

The data collected by the radar array during a Coherent Pulse Interval (CPI) , which will be the input to the radar processing algorithm, is normally considered as organized in a structure known as the "CPI Datacube". For a radar array with N elements, which transmits M pulses per Coherent Pulse Interval, and collects signals at L different range gates, the CPI datacube is organized as shown in Figure 1.

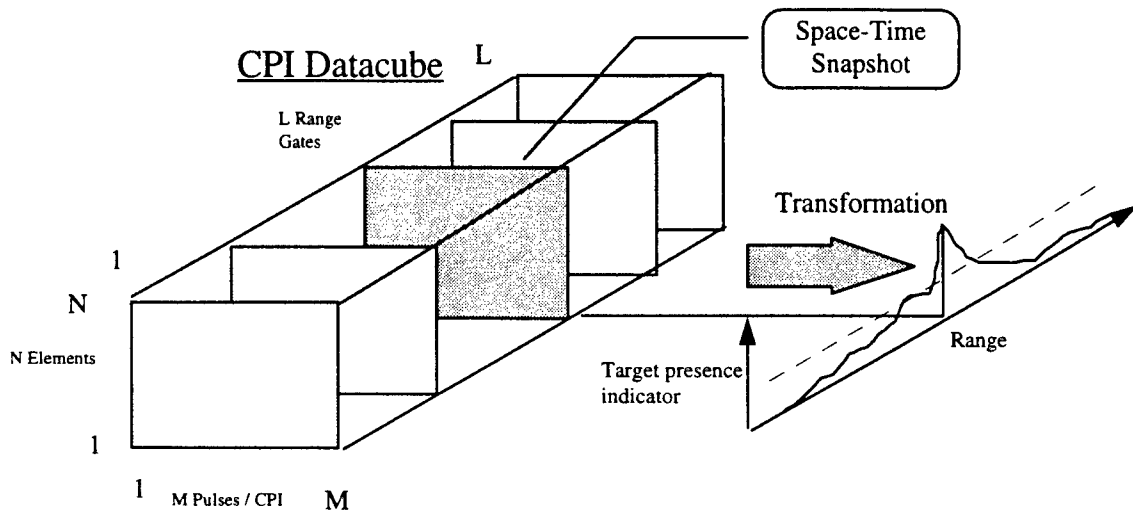


Figure 1. The CPI Datacube and the target detection process

The values associated with a single range gate are indicative of the scatterers at a single distance (range) from the radar instrument. This "slice" of the CPI Datacube is called a "Space-Time Snapshot". Figure 1 indicates that, ultimately, the objective of the data processing is the transformation of each space-time snapshot to evaluate some target parameter indicator and determine the presence of a valid target at a given range, to identify its azimuth and relative velocity. For this whole operation to be successful, the evaluation of each space-time snapshot must be really representative of the target presence alone. If the data in the space-time snapshot contains interference components (clutter and / or jammer) the target detection process will be misled.

Evidently, the processing applied to the received signals must attenuate or eliminate the interference and emphasize valid target components in each space-time snapshot. This type of filtering is in fact possible due to the different space-time characteristics of jammer and clutter, with respect to the target [1] [16].

This filtering has been approached through the implementation of a “space-time processor”, which is a linear filter that combines all the samples of the range gate of interest to produce a scalar output [17] :

$$z = w^H \chi \quad (1)$$

where χ is a the space-time snapshot vector (MN-by-1) for the range gate of interest, and w is an MN-by-1 weight vector.

In turn, an appropriate weight vector w , that will attenuate the interference and emphasize the target information is found as:

$$w = R_u^{-1} v_t \quad (2)$$

where v_t is the space-time steering vector for the target (which needs to be proposed repeatedly to investigate several possible azimuth and velocities of the target) and R_u is the interference covariance matrix:

$$R_u = E\{\chi_u \chi_u^H\} \quad (3)$$

Here χ_u represents snapshots ideally formed with interference data only.

In the filtering approach of equation (eq. 1) the removal of interference contributions to the output indicator z is effected through w , which, in turn depends on the data used to calculate R_u . This makes the space-time processor data-adaptive, which enables it to cope with the changing interference scenarios that the radar instrument may face. This is the basis of the “Space-Time Adaptive Processors” or STAP for radar. On the other hand, the required estimation of R_u and continuous upgrade of the weight vector makes this “fully adaptive STAP” method very computationally demanding, and susceptible to errors in the estimation of R_u .

The effect of the estimation of the interference covariance matrix (eq. 3) and its inversion to partially determine the weights of the space-time processor (eq. 2) amount to the identification of the interference components in the data and their cancellation. This is often emphasized by the display of the two-

dimensional Discrete Fourier Transform (DFT) of both, the original space-time snapshot showing the interference overpowering the target signal, and the “adapted STAP weights”, which shows nulls in the regions of that Doppler-Azimuth Spectrum (ADS) where the interference lies. It should be noted then, that the STAP approach relies on the identification of the interference components to achieve their cancellation.

We have proposed a complementary approach to the STAP method, in which we seek to identify the signature of a valid moving target in the Azimuth-Doppler Spectrum and isolate it from the rest of its components. In order to achieve the isolation of the target information we propose that the ADS signature of a valid moving target is determined by its relative velocity and angle, which defines its position in an abstract “Target Field”, t , and by a certain two-dimensional Impulse Response, H , which is determined by the number of elements and pulses per CPI of the radar array. The details of our deconvolution approach to target characterization are given in the next section.

Methodology

Figure 2a) illustrates the Azimuth-Doppler Spectrum (ADS) resulting of the simulation of a moving target at an azimuth of -30° and at a relative speed that is indicated by a Doppler shift of one forth of the Pulse Repetition Frequency, (PRF/4). Figure 2b) proposes the existence of a two-dimensional “Target Field”, where the presence of a moving target is indicated by a two-dimensional delta function at the appropriate coordinates of azimuth (x axis) and relative velocity (y axis). It should be noted that if the target were simulated at different azimuth and relative speed the ADS would still show the same type of signature, with a main lobe representing the target Doppler and Azimuth and sidelobes extending in both the ADS dimensions. Similar to the case in Figure 2, that new ADS for the target would be matched by a single two-dimensional impulse at the target relative velocity and angle in the target field. Observations like these prompted us to propose that the Doppler-Azimuth Spectrum expression of a valid moving target can be modeled as a process of two-dimensional, circular convolution between the abstract target field and a basic target “kernel” or impulse response determined by the M and N parameters of the radar system. Figure 3 illustrates this idea for a case where $N = M = 6$.

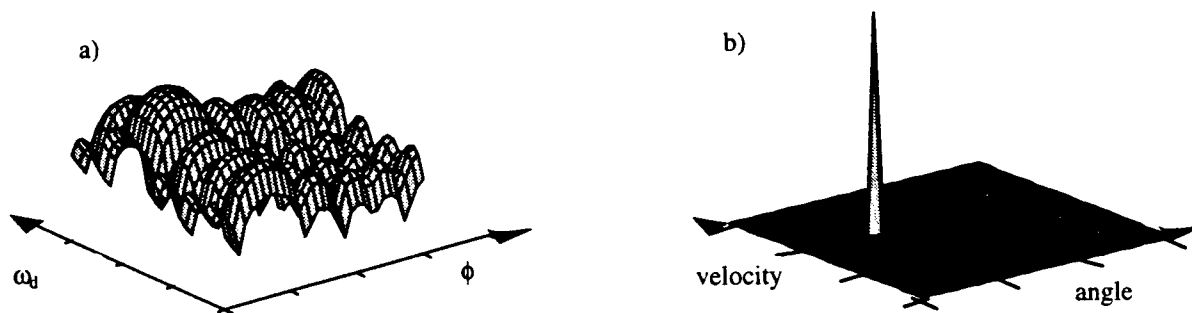


Figure 2. A moving target in a) the Azimuth-Doppler Spectrum, and b) the abstract "Target Field".

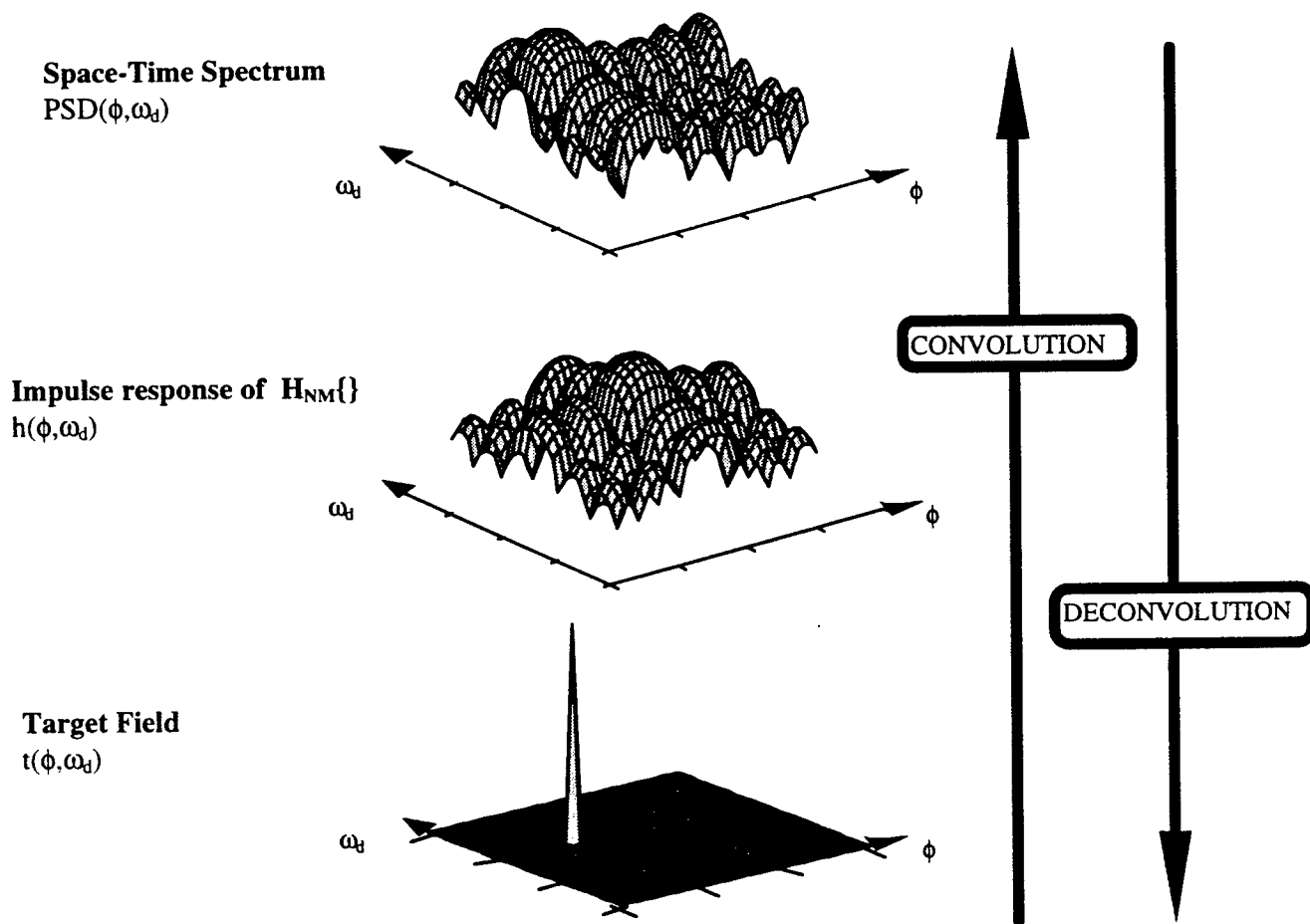


Figure 3. Relationship between the Target Field, The $H_{NM}\{\}$ transformation and the space-time spectrum under the convolution model.

In this process, it should be noted that the “kernel” or two-dimensional impulse response is, in fact, the Azimuth-Doppler Spectrum corresponding to an elementary target at zero relative velocity and zero azimuth. It is relevant to emphasize that the convolution model proposed for the generation of the spectrum is being based on the spectral expression of an elementary target.

If the existence of a valid target in the target field is reflected in the observed Azimuth-Doppler Spectrum (ADS) through two-dimensional, circular convolution with the known impulse response, H , it may be possible to revert the process and retrieve corresponding target field, indicating the relative velocity and angle of the target, through deconvolution. Furthermore, since the convolution model proposed is based on the kernel that projects an elementary target to the Azimuth-Doppler Spectrum, the associated deconvolution process will preferentially retrieve the spectral components due to targets, minimizing the retrieval of other spectral components that have a different structure, such as clutter.

To explore the possibilities of deconvolution for target characterization, we first express the discrete convolution process mathematically, as a matrix equation. In the context of discrete target fields and discrete Doppler-Azimuth Spectra, both of them with finite support, the circular convolution operation can be put in terms of a matrix product of the row-scanned target field, pre-multiplied by an appropriate matrix $H[i,j]$ to result in a vector representing the row-scanned Azimuth-Doppler Spectrum, or PSD:

$$\mathbf{H}_{(K^2 \times K^2)} \mathbf{t}_{(K^2 \times 1)} = \mathbf{PSD}_{(K^2 \times 1)} \quad (4)$$

In this equation K is the assumed number of discrete values for both dimensions of the Target Field and the space-time spectrum. The $[i,j]$ entries of the matrices involved in the above equation represent the following:

$\mathbf{t}[p(K-1)+q,1]$: the q^{th} element of the p^{th} row of the original K by K target field matrix

$\mathbf{PSD}[p(K-1)+q,1]$: the q^{th} element of the p^{th} row of the original K by K ADS matrix

$\mathbf{H}[i,j]$: the contribution to the PSD in the azimuth,Doppler combination represented by the $[i,1]$ element of the vector \mathbf{PSD} , due to an assumed elementary target of unit power located at the azimuth, Doppler combination represented by the $[j,1]$ element of the vector \mathbf{t} .

Expressing the convolution process in the form of equation (4) makes it possible to approach the deconvolution process with two types of methods. Initially, changing the known terms of equation (4) to be H and PSD, with t as the unknown vector, transforms the equation into a set of K^2 simultaneous linear equations. A direct solution, if one exists, is likely to include elements of t that are negative. This would not provide a valid solution for the target field, since we have proposed that its elements represent the presence ($t > 0$) or absence ($t = 0$) of a target at the corresponding Doppler and Azimuth. Instead, the minimization of the norm $\|Ht - PSD\|^2$, subject to the constraint that all the elements of t must be positive or zero can be sought.

Alternatively, equation (eq. 4) can be used in iterative methods that seek to minimize an error or residual between the observed PSD and an estimation obtained by convolving a proposed target field with the known kernel. In these methods the residual is progressively minimized by refining the target field estimate using information from the residual obtained in the previous iteration. In these methods the positivity of the target field result is promoted by applying a positivity constraint operator on the current target field estimate at each iteration:

$$C[t(n_1, n_2)] = \begin{cases} t(n_1, n_2) & \text{if } P1 \leq n1 \leq Q1, P2 \leq n2 \leq Q2 \\ & \text{and } t(n_1, n_2) \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

The mechanisms for two-dimensional deconvolution that were explored in these project were as follows:

- **The Non-Negative Least Squares (NNLS) method:**

The Non-Negative Least Squares (NNLS) minimization algorithm of Lawson and Hanson [6] implements the kind of norm minimization required to solve equation (eq. 4) for the target field, t.

$$x = \min_t \|Ht - PSD\|^2 \quad \text{such that } x \geq 0 \quad (6)$$

It develops recurrent solutions of increasing order, checking at each order that the solution contains only values above a certain tolerance level (close to zero). If the solution at any particular order is not completely non-negative it is regenerated until it satisfies the constraint. The non-negative solutions continue to be generated until they reach the true order of the problem.

The NNLS deconvolution approach was the one used during the Summer Faculty Research Program project preceding this SREP project. It has been found to result in recovered target fields that are comprised by delta-like components, which conforms to our assumptions for the target field. This method, however, is susceptible to reproduce small and isolated target field components as a response to irregularities in the Azimuth -Doppler Spectrum. Additionally, although its convergence in a finite number of steps is guaranteed [6], the time required by this method is data-dependent and may be longer than others.

• **The Richardson-Lucy Methods for iterative deconvolution:**

The fundamental aspects of this iterative deconvolution method were developed independently by Richardson [10] and Lucy [7] independently. Their approaches have been grouped together because they are both derived from Bayes theorem, both treat the Target Field $t(i, j)$, Impulse Response $H(i, j)$, and Doppler-Azimuth Spectrum $PSD(i, j)$ as probability frequency functions. The only difference in their application is the method of iteration. Richardson's method is directly presented as six nested loop summations (four are shown below, the two missing are needed for running indices i and j) and an outer loop for each iteration, r [10]. The nested loops make this approach extremely time consuming.

$$t_{r+1}(i, j) = t_r(i, j) \sum_{m=i}^e \sum_{n=j}^f \frac{PSD(m, n) H(m - i + 1, n - j + 1)}{\sum_{p=a}^b \sum_{q=c}^d t_r(p, q) H(m - p + 1, n - q + 1)} \quad (7)$$

$$t_1(i, j) = \sum_{m=i}^e \sum_{n=j}^f \frac{PSD(m, n) H(m - i + 1, n - j + 1)}{\sum_{p=a}^b \sum_{q=c}^d H(m - p + 1, n - q + 1)} \quad (8)$$

Where:

$$a = (1, m - K + 1)_{\max} \quad b = (m, I)_{\min}$$

$$c = (1, n - L + 1)_{\max} \quad d = (n, J)_{\min}$$

$$e = i + K - 1 \quad f = j + L - 1$$

$$i = \{1, I\} \quad j = \{1, J\}$$

K, L are the dimensions of $H_{k,l}$

I, J are the dimensions of $t_{i,j}$

The slow execution of this method and the unimpressive results obtained through it in preliminary attempts to deconvolve the Azimuth-Doppler Spectrum prompted us not to continue investigating it.

On the other hand, Lucy's method is implemented with just one loop executing the iterations and a series of matrix multiplications, where \otimes indicates circular convolution, and PSD is the original ("observed") PSD , while PSD_r is the estimation of the spectrum at iteration r [7].

$$t_{r+1} = t_r \left(\frac{PSD}{PSD_r} \otimes H^T \right) \quad (9)$$

$$PSD_r = H \otimes t_r \quad (10)$$

The circular convolution is, in practice, implemented by applying the current target field and PSD vectors to equation (eq. 4).

This method was significantly faster than Richardson's. On the other hand, when applied to the deconvolution of simple simulated Doppler-Azimuth Spectra, Lucy's method did not appear to synthesize a target field with clear indications for the azimuth and Doppler of the targets.

- **Van Cittert's Method for Iterative Deconvolution:**

This is a very simple iterative deconvolution method that strives to minimize the residual by feeding it back to upgrade the target field estimate:

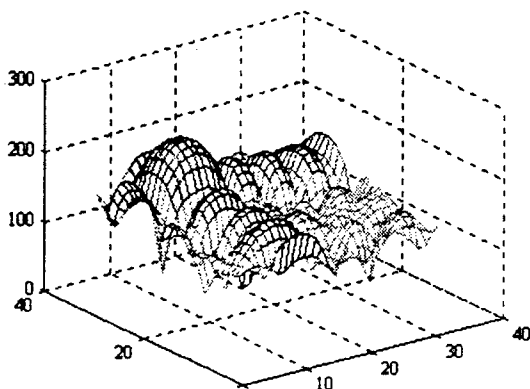
$$t_{r+1}(i, j) = t_r(i, j) + \alpha R_r(i, j) \quad (11)$$

$$R_r(i, j) = PSD(i, j) - H \otimes t_r(i, j) \quad (12)$$

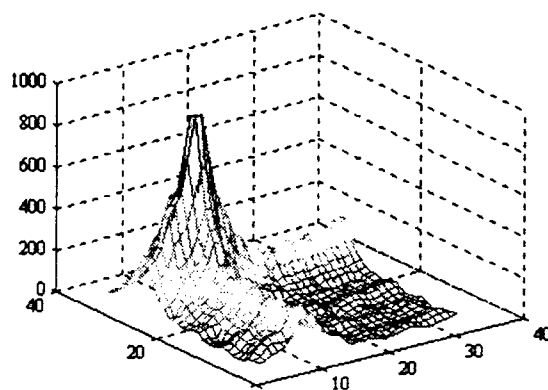
This process iterates, either for a given number of iterations, r_{MAX} or until the data in the residual matrix is considered negligible[15].

This method allows control over the number of iterations to apply, which determines the total number of calculations to execute. We have found that 100 iterations is normally an appropriate number for the deconvolution of Azimuth-Doppler spectra. Under these conditions Van Cittert's method is one of the fastest means to execute the deconvolution of the ADS. On the other hand, this method normally results in target field components that are less delta-like than those from NNLS deconvolution. Figure 4 shows the

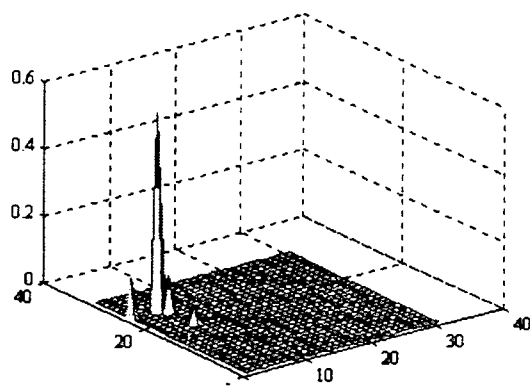
target fields resulting from the deconvolution of the same Doppler-Azimuth Spectrum, through the NNLS, Lucy's and Van Cittert's deconvolution methods. Because of their characteristics of accuracy and speed, the NNLS and Van Cittert's methods were chosen for further study.



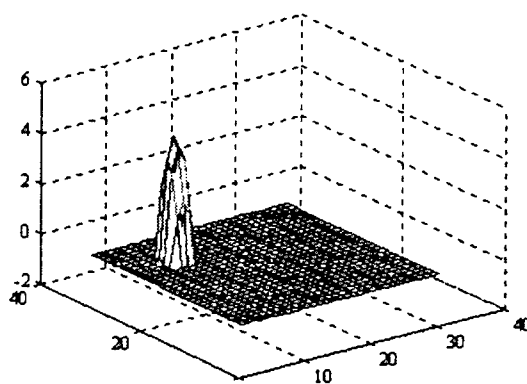
a) Azimuth-Doppler Spectrum for a single target



b) Deconvolution of a), through Lucy's Method



c) Deconvolution of a) through the NNLS method



d) Deconvolution of a) through Van Cittert's method

Figure 4. The Azimuth-Doppler Spectrum of a single elementary target without interference and the results of its deconvolution through three different methods.

In investigating the performance of the different deconvolution methods, it was found that the NNLS approach yielded the most delta-like target indicators. However, the processing time for NNLS tends to be longer than for other methods and it is, in fact, data-dependent. The dependency of the NNLS computation time on the specific data being processed is due to the varying need to reject solutions with negative or very small components, when they are tested and do not surpass a given tolerance level, preset in the NNLS algorithm. Initially, it was hoped that relaxing the tolerance parameter in the algorithm could accelerate the convergence, even if this resulted in the appearance of small negative target field components, which would be neglected.

Unfortunately the variation of the tolerance parameter over a wide range of values did not result in any appreciable improvement of the speed for NNLS. The following table lists the computation time for the deconvolution of a 31-by-31 Azimuth-Doppler Spectrum, using a tolerance parameter that was modified by the factors shown from the default value assigned by Matlab ®. This value is determined by Matlab ® as

$$tolerance = \max(\text{size}(A)) * \text{norm}(A,1) * \text{eps} \quad (13)$$

where *eps* is a machine-dependent accuracy parameter that represents the distance between 1 and the next largest floating-point number. The tests were run in a 200 Mhz R4400-based SGI Challenge machine, for which *eps* = 2.2204e-016.

Tolerance Factor	Comput. Time (min)	Variable name
0.0001	0.90421	nctf_0001
0.001	0.94051	nctf_001
0.01	0.94018	nctf_01
0.1	0.9368	nctf_1
1	0.93762	nctf1
10	0.93963	nctf_10
100	0.9395	nctf_100
1000	0.93834	nctf_1000

Table 1. NNLS Computation time for different Tolerance Factors, on the same ADS for 1 simulated target.

As observed in Table I, the variation of the default NNLS tolerance by factors ranging from 0.001 to 1000 did not cause any significant reduction in the execution speed, for this particular problem. The recovered target field was very similar in all cases and resembled the one NNLS result shown in Figure 4.

The convolution model proposed for the transition from a target field to the corresponding Azimuth-Doppler Spectrum was based on the assumption of an “impulse response” or kernel, which specifically resulted in the spectral expression of a valid target in the ADS. The capability of the deconvolution method to minimize the impact of other spectral components depends on the difference of the spectral structure of interference elements with respect to the structure of a valid target. The following figures display Azimuth-Doppler Spectra obtained from the “MountainTop Toolbox” created by Yaron Seliktar from The Georgia Institute of Technology [11]. This custom Matlab ® toolbox provides a program and a front-end for the simulation of multi-element radar array signals that incorporate returns from a target, clutter and jammer, allowing the user to specify the characteristics of the simulated components and the operating parameters of the radar. (All the simulated data for this project were generated with the “MountainTop Toolbox”, slightly customized for our purposes).

Figure 5 represents the power signature of an elementary target simulated at 0 dBs, at an angle of 20 degrees with respect to the array broadside and with a Doppler frequency of -200 Hz. Figure 6 represents the ADS for simulated clutter at 40 dBs, assuming a platform speed of 100 m/s. Finally, Figure 7 represents simulated barrage jammer at 0dBs and from an azimuth of 30 degrees with respect to the broadside of the array. It should be noted that, in fact, the isolated target appears represented in the ADS by a pattern similar to the ones considered before, except that reflecting the 16 pulses per CPI ($M=16$) and the 14 array elements ($N = 14$) assumed in the simulation. On the other hand, the spectral structure of clutter displays the characteristic “clutter ridge”, which is, in fact, very different from the spectral structure of a target. From this observation, we can expect to apply the process of deconvolution to retrieve target components preferentially over clutter components.

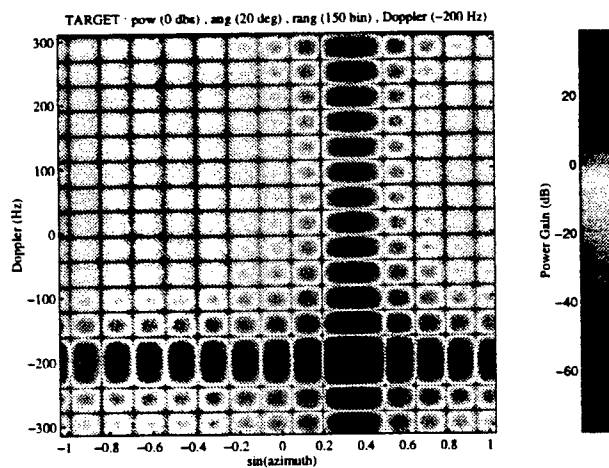


Figure 5. Azimuth -Doppler Spectrum of a target (0 dB) at 20° Azimuth and -200 Hz Doppler

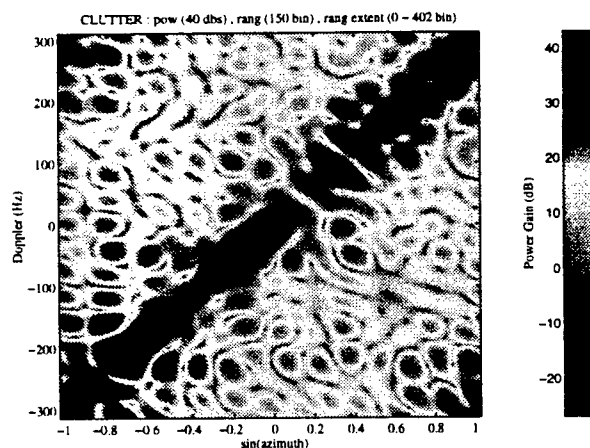


Figure 6. Azimuth -Doppler Spectrum of clutter (0 dB) for a platform velocity of 100 m/s

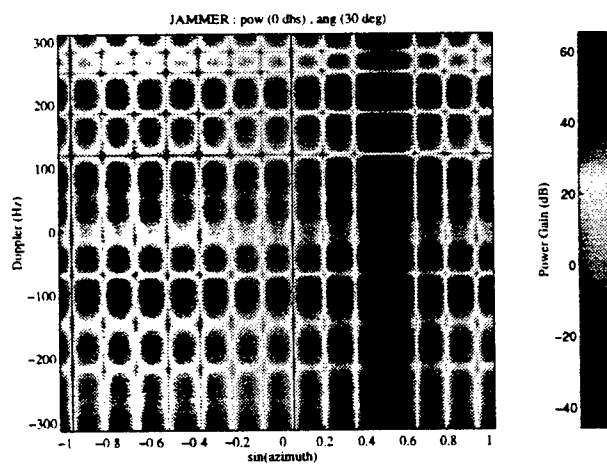


Figure 7. Azimuth -Doppler Spectrum of jamming interference (0 dB) from 30° Azimuth

While the spectral structure of the clutter interference is, in general, different from that of a target, the jammer interference appears in the ADS as a high ridge at a constant azimuth, spanning all the Doppler frequencies. Unfortunately, we have found in our preliminary work with ADS deconvolution that the jammer spectrum is, in fact, retrieved as a series of components in the target field, all lined up at the same azimuth. To prevent this effect from dominating the recovered target field, we found it necessary to apply a pre-processing scheme to remove the jammer components before submitting the ADS to the deconvolution operation. In this way the pre-processor will attenuate the jammer interference first and then the deconvolution process will preferentially retrieve the true target components of the ADS.

Attenuation of the jammer interference only can be accomplished by a process that filters in the spatial dimension only. This means that the pre-processor acts only on data collected simultaneously by all the array elements. Since the azimuth of incidence of the jammer will very likely change through time, the spatial "notch" imposed by the desired pre-processor for its cancellation must also be defined in a data-adaptive manner. In fact, the equations that govern the performance of an adaptive spatial filter for jammer cancellation will have the same structure as equations (eq. 1), (eq. 2) and (eq. 3), except that this time they manipulate a vector with only the samples collected by the N array elements at a single sampling time, the spatial snapshot : X .

Here the output indicator for the spatial-only filter would be:

$$z_s = w^H X \quad (14)$$

In this case, the N weights needed for the spatial filter are also based on the interference covariance matrix, R_s and a target spatial steering vector v_s :

$$w = R_s^{-1} v_s \quad (15)$$

Fortunately, the interference covariance matrix for the space-only filter is estimated from the spatial snapshot only, of size N [2] [3]:

$$R_s = E\{XX^H\} \quad (16)$$

In practice the expectation operator in the above equation is implemented as a time-average:

$$R_s \approx \frac{1}{K} \sum_{i=1}^K X_i X_i^H \quad (17)$$

This estimate of the spatial interference covariance matrix would next have to be inverted. This yields the name "Sample Matrix Inversion", or SMI, for this form of adaptive filtering. In practice, however, it has been found that, in order to improve the condition number of the estimated R_s , it is advantageous to add a constant term, β , to all the elements in its diagonal. This method is called Sample Matrix Inversion "with Diagonal Loading" and has been shown to also yield an antenna pattern with smaller side lobes [2] [3].

In our pre-processor, we applied SMI with Diagonal Loading ($\beta = 0.01$) to generate an estimate of R_s^{-1} . Then this matrix was applied to the data in the CPI under study without using a spatial steering vector. If we isolate the effect of the steering vector in the determination of the output indicator z_s , we realize that an equivalent vector, X' , from which the interference has been removed, without emphasizing a specific "look-direction" can be found:

$$z_s = w^H X = (R_s^{-1} v_s)^H X = v_s^H ((R_s^{-1})^H X) = v_s^H (X^H R_s^{-1})^H = v_s^H X'^H \quad (18)$$

where

$$X' = X^H R_s^{-1} \quad (19)$$

is the spatially-filtered spatial snapshot.

In the experimental cases where a jamming component was simulated or known to exist, we used the above procedure as a pre-processing stage, substituting all the spatial-snapshots X in the CPI under study by the corresponding filtered X' .

Results

The ability of the deconvolution process to preferentially retrieve target information in the presence of interference components was demonstrated by its application to radar data of several types. In this report some cases representative of the following categories are shown:

- a) Synthetic data (generated through the "MountainTop Toolbox" [11]), including the simulation of target and clutter.

In this case the deconvolution approaches were directly applied to the simulated data.

- b) Synthetic data including the simulation of a target, clutter and jamming interference.

In this case the "SMI with diagonal loading" approach was used first, as a pre-processing stage, and then the deconvolution method was applied.

- c) Real data collected through the DoD "Mountaintop Program" [13][14], and retrieved from the Common Research Environment for STAP Technology (CREST) Datalibrary.

In all cases both the NNLS and Van Cittert's methods for constrained deconvolution were applied, for comparison.

a) Deconvolution of synthetic data with simulated clutter and target.

During the Summer Faculty Research Program, the possibility of preferentially retrieving the target ADS spectral components, over the components due to clutter was first examined in simulated cases with ADS at resolution, $K = 31$. (That is, the ADS was represented by a matrix of size 31×31). This was needed since the interpreted Matlab ® program implementing NNLS deconvolution was relatively slow and using larger resolutions would have made the processing time too long. Within those constraints, the simulations were carried out assuming 6 antenna elements ($N=6$) and 6 pulses per CPI ($M = 6$). The two-dimensional impulse response (or "kernel") for these simulation conditions is shown in Figure 8. In that preliminary study only the NNLS deconvolution method was used, with encouraging results. Figure 9 shows a 31×31 synthetic ADS where a target was simulated with a power of 0 dB, at -30° Azimuth and 150 Hz Doppler. This same ADS includes a simulated clutter component at 30 dB power, for which the platform velocity has been assumed at 100 m/s. This clutter ridge can be seen distorting the spectral pattern of the target along a diagonal "ridge", as it is typical for this form of interference.

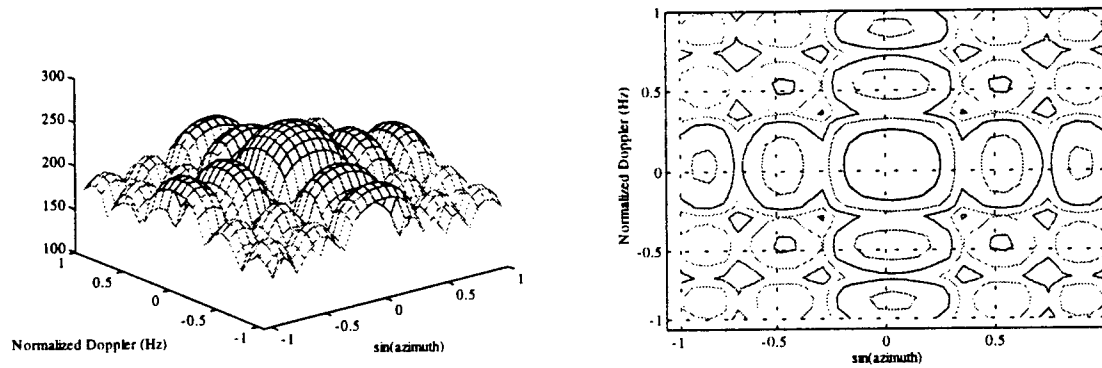


Figure 8. 2D- Impulse Response for a system with $N=M=6$, $K=31$

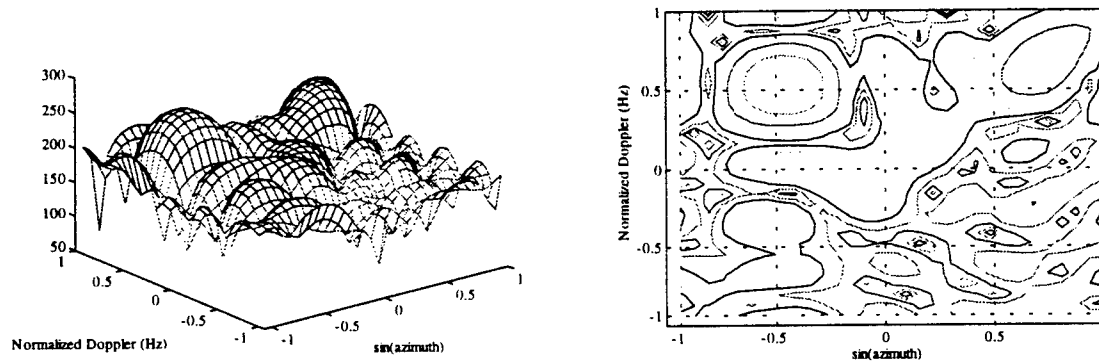


Figure 9. ADS for a simulated target (0 dB, -30° , 150 Hz) and clutter (30 dB, 100 m/s), $K=31$

The results obtained by applying NNLS deconvolution to this ADS are shown below, in Figure 10. The retrieved target field is almost uniquely defined by a delta function at the velocity (Doppler) and angle (azimuth) at which the target was simulated, except for some small adjacent peaks. It is believed that those small spurious components result from the slight configuration mismatch between the spectrum of the target and the kernel used for the deconvolution, which is, in turn, a result of the limited resolution of the ADS. The clutter ridge, however, has not been represented in the target field.

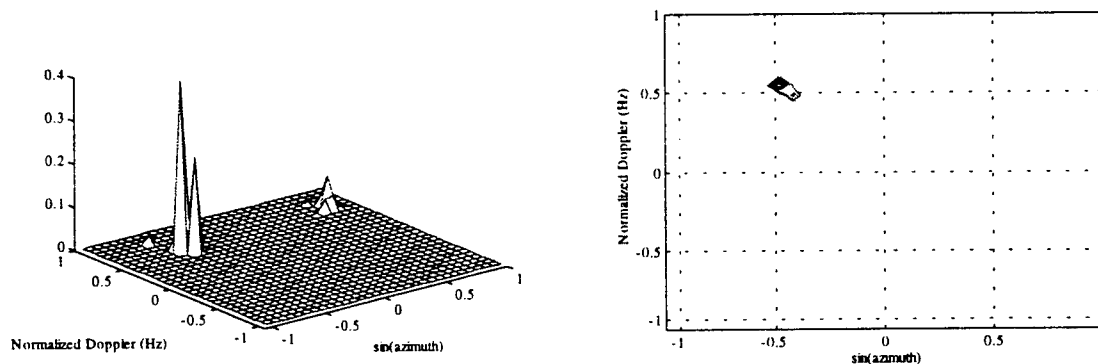


Figure 10. Target Field recovered from the ADS in Figure 9, through NNLS

The results of applying Van Cittert's iterative constrained deconvolution over 100 iterations, with a rate factor $\alpha = 1 \times 10^{-6}$, to the same ADS are shown in Figure 11. These results indicate that the residual minimization of Van Cittert's method yields a target indication that is not a single delta function. Instead, the target field converges to a target representation that spans a few cells in the target field, around the target characteristics (azimuth, Doppler). This feature of the Van Cittert's method was verified through subsequent applications to other data. Although, in principle, this type of result is not as specific as the one potentially offered by NNLS, Van Cittert's method was found to be generally faster than NNLS deconvolution. For the particular example considered above, Van Cittert's deconvolution took less than one tenth of the time required by NNLS deconvolution. In presenting this timing comparison, it must be kept in mind that Van Cittert's timing is completely defined by the size of the ADS matrix, K , and the number of iterations allowed. On the other hand NNLS timing is influenced by the size of the ADS matrix, but then it depends significantly on the nature of the data itself, which determines how many times the inner loop in the algorithm needs to be repeated for each progressively bigger solution order.

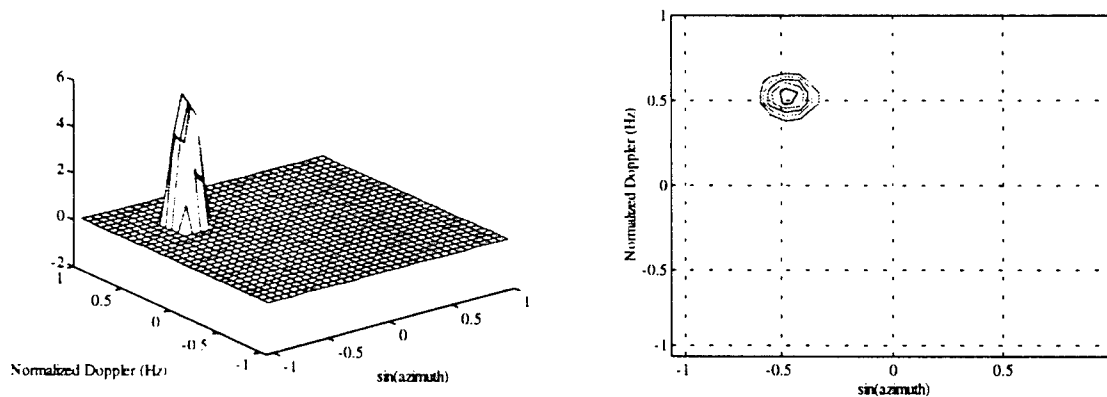


Figure 11. Target Field recovered from the ADS in Figure 9, through Van Cittert's method

When the power level of the synthetic clutter was raised to 40 dB, some of the concentrated clutter patches reached levels that made them resemble the target power configuration. This caused the retrieval of spurious peaks in the recovered target fields, in addition to the ones associated with the true target. From the results shown in the figures below, it can be realized that NNLS deconvolution is more susceptible to these effects than Van Cittert's. The latter, in fact, only reflected a significant spurious peak at the corner of the ADS where the clutter ridge ends. The structure of the ADS in this area shows, in effect, a partial main lobe with smaller lobes around it, which are perceived by the deconvolution methods as its side lobes.

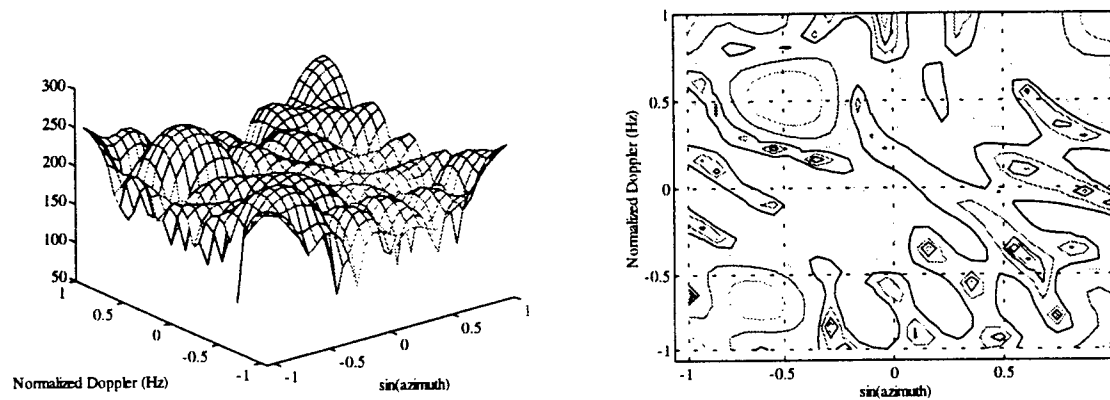


Figure 12. ADS for a simulated target(0 dB, -30° , 150 Hz) and clutter (40 dB, 100 m/s), $K = 31$

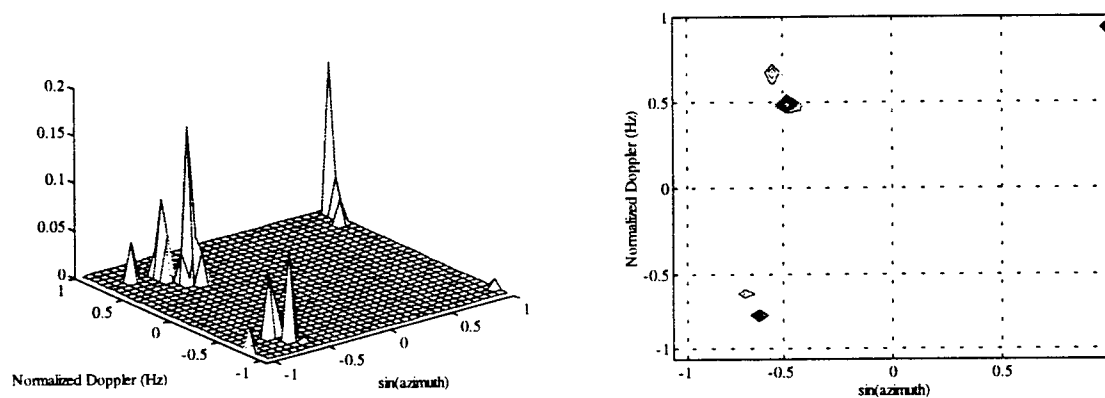


Figure 13. Target Field recovered from the ADS in Figure 12, through NNLS

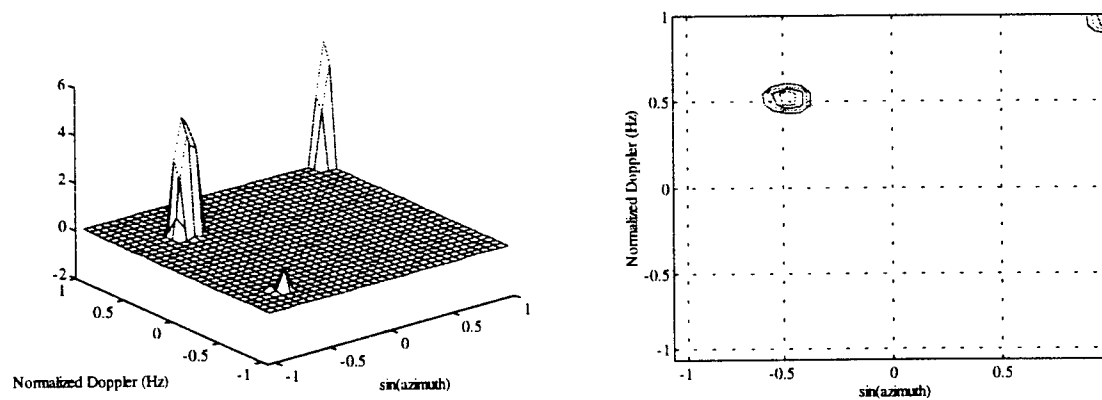


Figure 14. Target Field recovered from the ADS in Figure 12, through Van Cittert's method

Once both deconvolution methods were implemented in (compiled) ANSI C, it was possible to experiment with better ASD resolutions, at a reasonable execution time. This step was necessary in order to perform deconvolution on Azimuth-Doppler spectra simulated with more realistic number of elements, N , and pulses per CPI, M . Our next level of experimentation was with ADS matrices of size 48×48 and N and M parameters which resemble those for the radar system used in the Mountaintop data, ($N=14$, $M=16$). The figure below shows the 48×48 ADS resulting from simulating the same scenario as before (one target at 0dB, -30° Azimuth, 150 Hz Doppler, clutter at 30 dB power, platform velocity 100 m/s) but assuming $N=14$ and $M=16$.

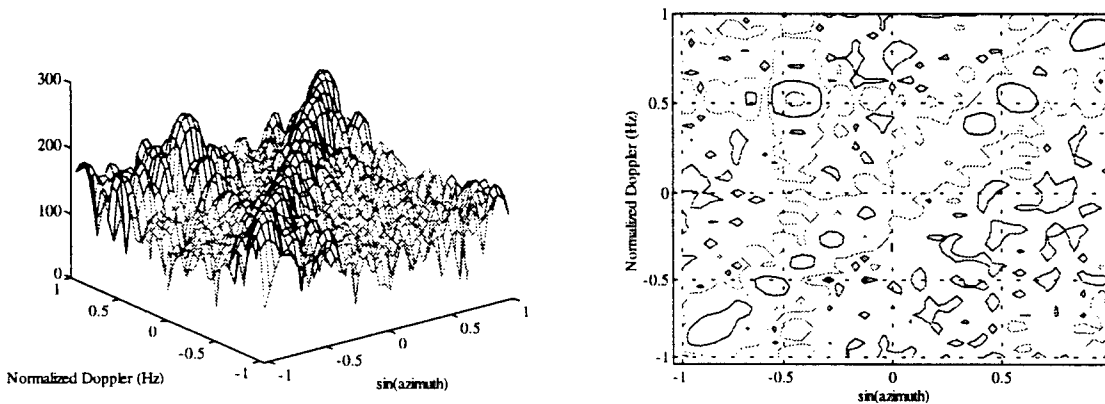


Figure 15. ADS for a simulated target(0 dB, -30° , 150 Hz) and clutter (30 dB, 100 m/s), $K=48$

The target field obtained through Van Cittert's deconvolution is shown below: The target information is recovered and the clutter components are virtually disregarded.

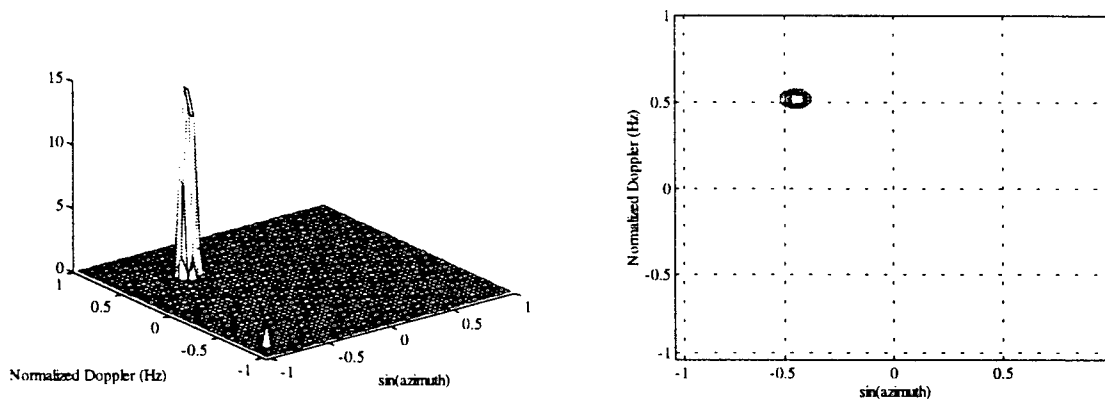


Figure 16. Target Field recovered from the ADS in Figure 15, through Van Cittert's Deconvolution

The results obtained through NNLS deconvolution are shown in the next figure. In addition to the target field components associated with the target, this result includes groups of smaller components at the ends of the clutter ridge and at the intersections of it with the side lobes of the target spectral expression. The power contributions of both target (side lobes) and clutter ridge are superimposed in those intersections, which makes that region of the ADS resemble a target main lobe. This effect is reinforced by the fact that the target side lobes and clutter ridge approximate the structure that the side lobes of a mainlobe in the intersection would have.

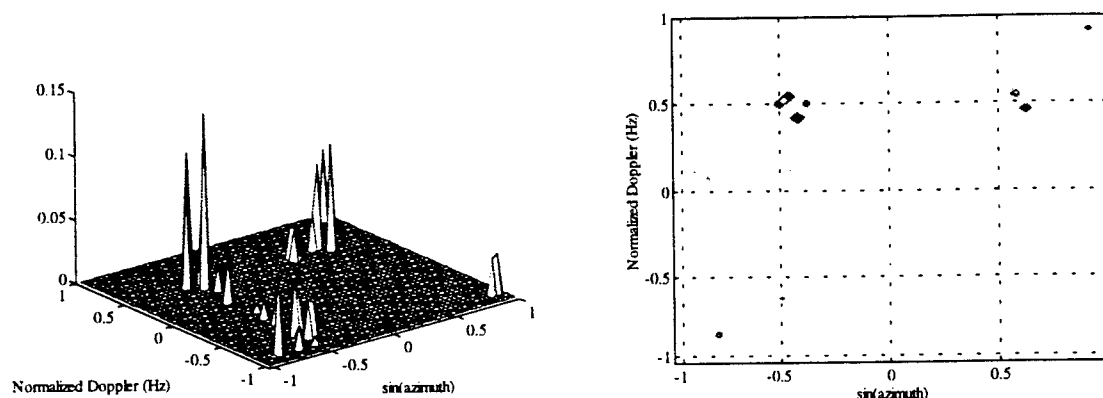


Figure 17. Target Field recovered from the ADS in Figure 15, through NNLS

b) Deconvolution of synthetic data with simulated target, clutter and jammer

To verify the efficiency of the proposed pre-processor for jamming attenuation, a data set was simulated containing a target at 0 dB, -30° azimuth, 150 Hz Doppler, along with clutter at 0 dB (platform velocity 100 m/s) and barrage jammer at 30 dB, from an azimuth of 30° . In this first case the attenuation of the jammer with low-power clutter only is studied.

The following two figures show the original simulated ADS and the ADS of the data after SMI spatial filtering, with covariance matrix diagonal loading. It should be noted that the jammer component, which overwhelms the original ADS has been effectively "notched out" after SMI spatial filtering. In this intermediate result the power representations of the target and the clutter are now detectable.

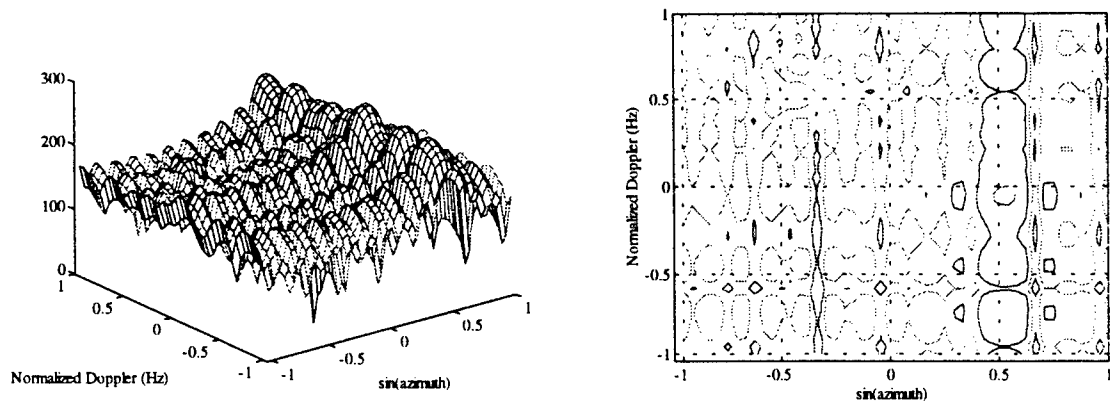


Figure 18. ADS for a simulated target(0 dB, -30° , 150 Hz), clutter (0 dB, 100 m/s) & jammer (30 dB, 30°)

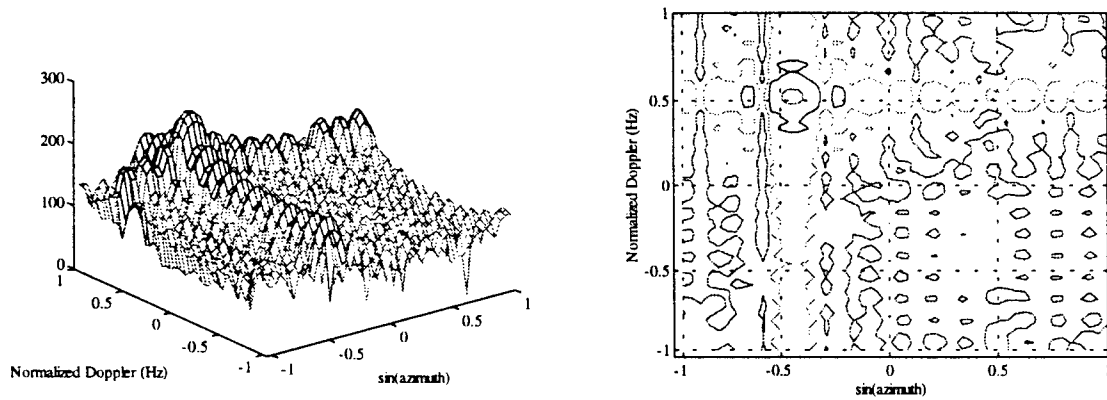


Figure 19. ADS of the data in Figure 18, after SMI jammer elimination

After the SMI pre-processor has removed the jammer interference NNLS and Van Cittert's deconvolution are applied to the intermediate result, to retrieve the target information:

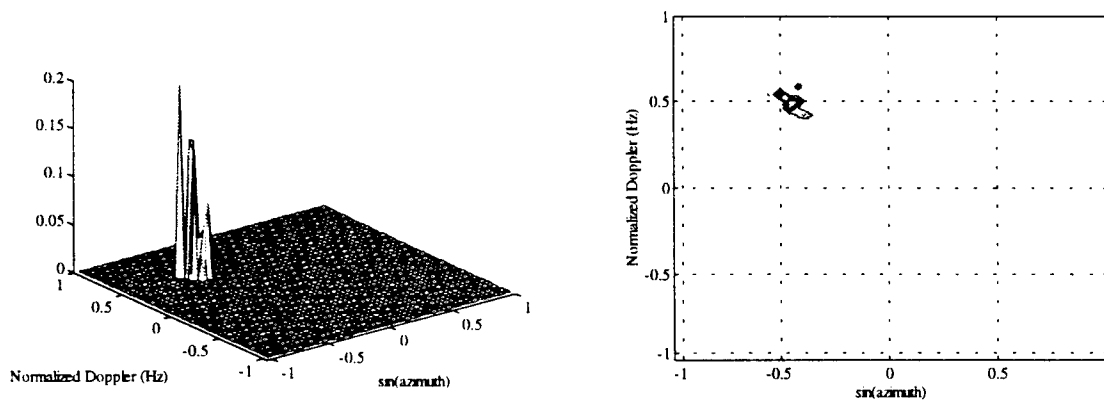


Figure 20. Target Field recovered from the ADS in Figure 19, through NNLS

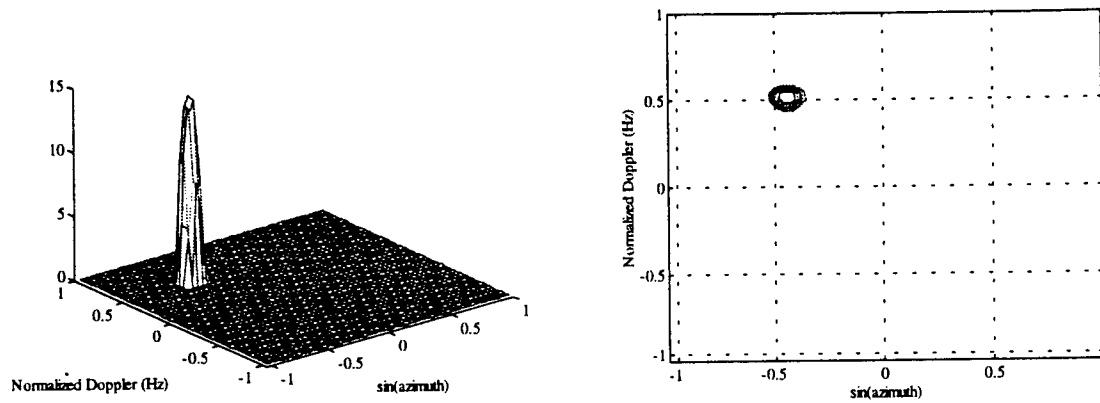


Figure 21. Target Field recovered from the ADS in Figure 19, through Van Cittert's Deconvolution

Both approaches are capable of retrieving the target component, NNLS generates some small peaks around the actual target characteristics (velocity and azimuth). Van Cittert's results retrieve only the information of the target, disregarding the clutter components.

The complete process involving the SMI pre-processor and the deconvolution approaches were tested also with a scenario similar to the previous one in which the power of the clutter has been increased to 30 dB. The original ADS, the intermediate result after the pre-processor and the final results obtained from Van Cittert's and NNLS are shown in the following figures.

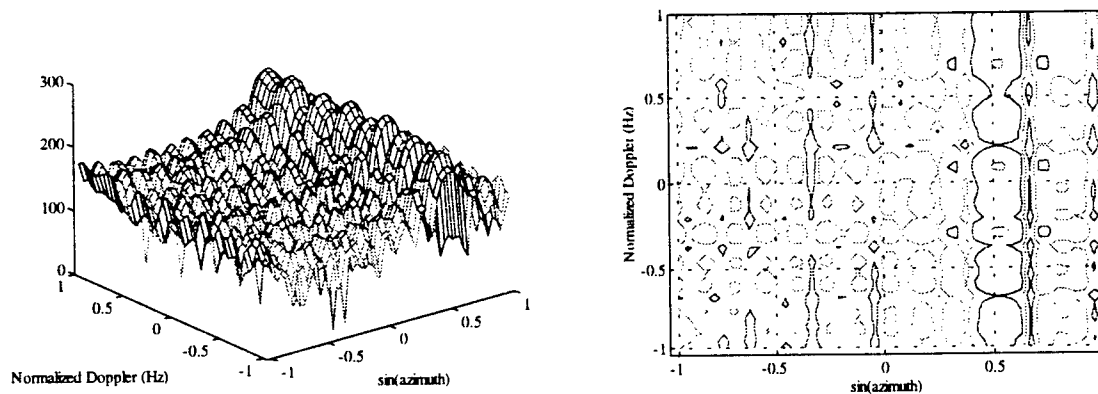


Figure 22.. ADS for simulated target(0 dB, -30° , 150 Hz), clutter (30 dB, 100 m/s) & jammer (30 dB, 30°)

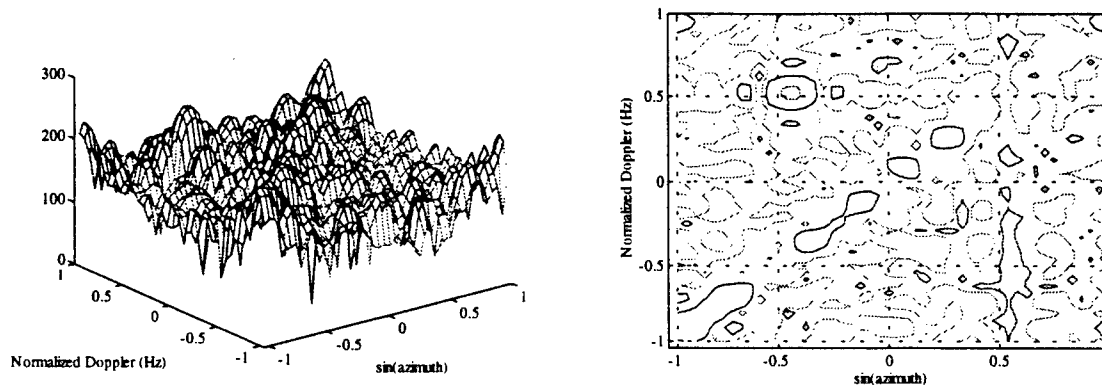


Figure 23. ADS of the data in Figure 22, after SMI jammer elimination

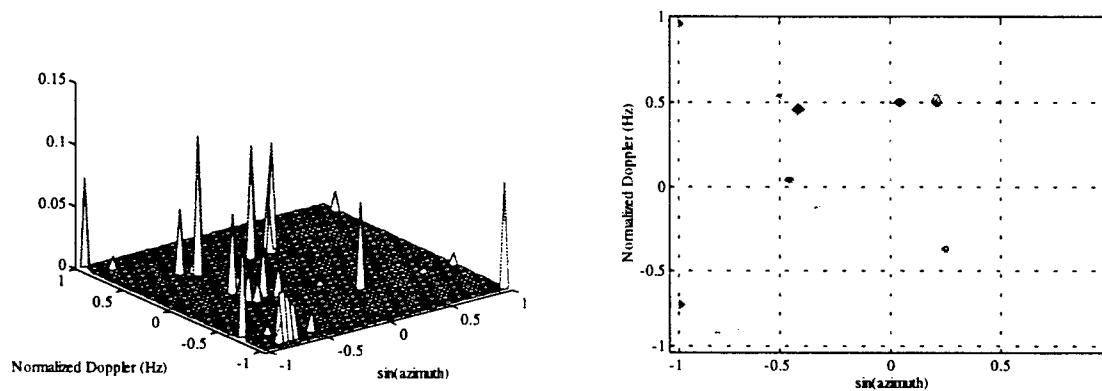


Figure 24. Target Field recovered from the ADS in Figure 23, through NNLS

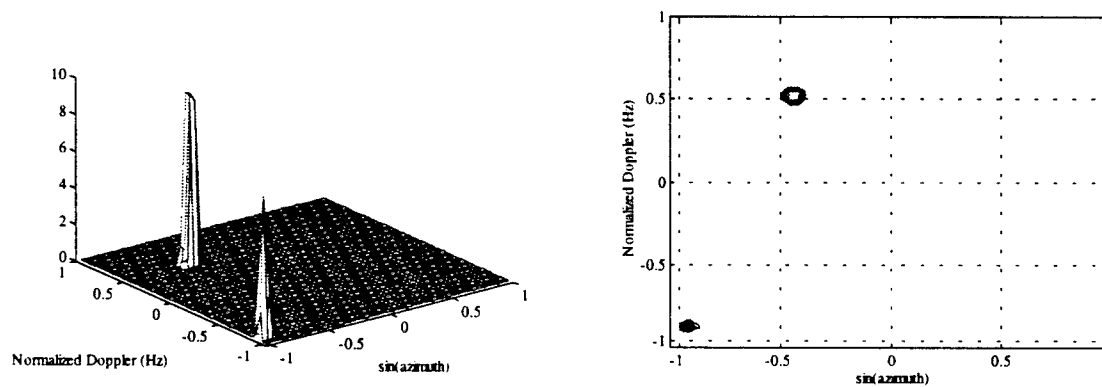


Figure 25. Target Field recovered from the ADS in Figure 23, through Van Cittert's Deconvolution

In this case Van Cittert's results indicate the characteristics of the target but also display a component that reflects the end of the clutter ridge, as explained before. On the other hand, the distortion introduced in the clutter components by the SMI spatial filtering affects NNLS more severely, causing it to represent small elements in the intersections of the target side lobes and the clutter ridges.

c) Deconvolution of real data collected through the DoD "Mountaintop Program", and retrieved from the Common Research Environment for STAP Technology (CREST) Datalibrary.

A number of files from the CREST Datalibrary were processed for target isolation. These files were collected with the Radar Surveillance Technology Experimental Radar (RSTER) system, employing 14 receiving elements ($N=14$) and with 16 pulses per CPI ($M=16$). The example presented here is the data in file "stap 1004v1.1030, cpi 3724, range bin 20". In this data set the Inverse Displaced Phase Center Array (IDPCA) technique was used to emulate the movement of an airborne platform with a stationary radar system, spreading the spectral manifestation of the clutter in the Doppler dimension [8]. In this real data set, however, the intensity of the clutter ridge is not uniform, with patches at about -20° and 60° azimuth being stronger scatterers. A moving target simulator (MTS) had been set up to emulate a target at about -5° and about 100 Hz Doppler, very close to one of the strong background scatterers. In fact the figure below shows how the clutter patch and the MTS are almost integrated in the original Azimuth-Doppler spectrum.

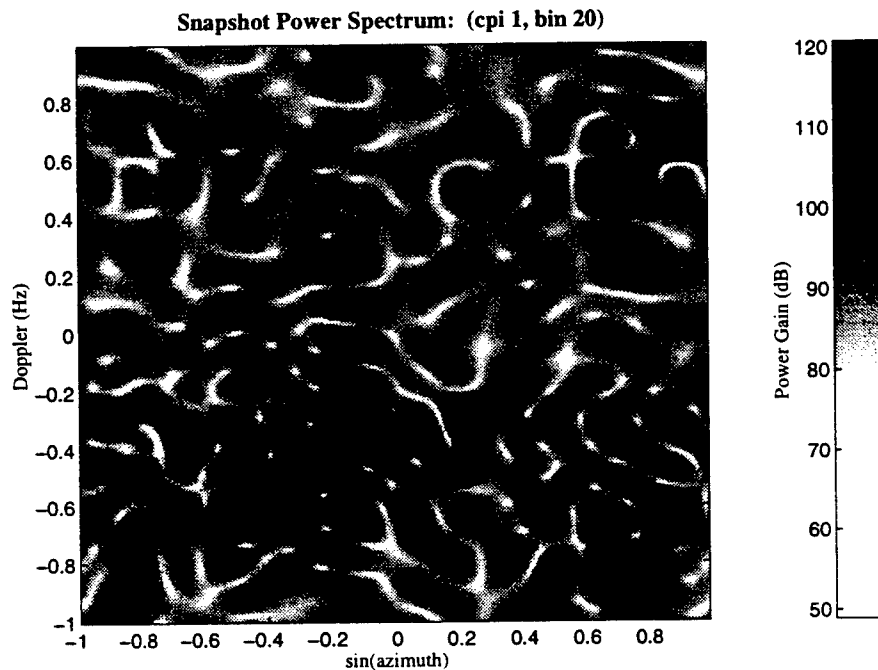


Figure 26. Azimuth-Doppler Spectrum for CREST file: STAP1004v1.1030.CPI3721, range bin 20

The following figures show the results of NNLS and Van Cittert's deconvolution of the ADS shown above. Van Cittert's deconvolution retrieves the target, showing a single smaller peak derived from the strong clutter patch, but in this target field both nonzero elements are clearly separated, and the target indication is predominant. NNLS generated a target indication but it also developed a number of smaller components, including one for the second patch of strong clutter at about 60° Azimuth.

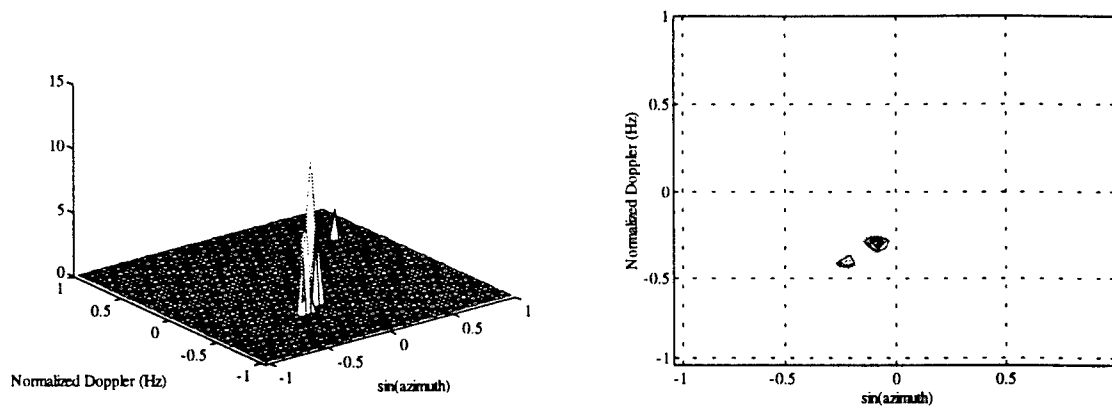


Figure 27. Target Field recovered from the ADS in Figure 26, through Van Cittert's Deconvolution

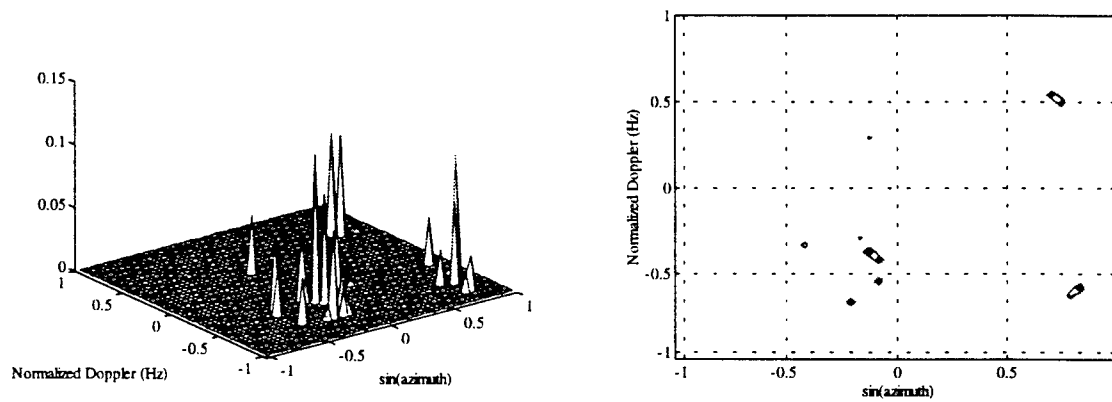


Figure 28. Target Field recovered from the ADS in Figure 26, through NNLS

Discussion

The previous section showed the application of the deconvolution process to progressively more realistic scenarios for the Azimuth-Doppler Spectrum of simulated and real array radar signals. In all cases the results obtained from two different forms of implementation, Van Cittert's and NNLS were shown and compared.

In the simplest case, in which idealized signal elements (target and clutter) were simulated at a low spectral resolution ($K = 31$), for a system with only 6 array elements and 6 pulses transmitted per CPI ($N = M = 6$), both methods were able to successfully retrieve the information of the target, disregarding the spectral components related to the clutter interference. In these cases the NNLS method of deconvolution seemed to provide a more specific characterization of the target, with respect to its Doppler and azimuth, since the recovered target fields would only involve one or two non-zero values, in each case. Van Cittert's method of iterative constrained deconvolution, on the other hand, resulted in target fields where a thin two-dimensional pulse would be observed, representing the target. Although this pulse was composed of more than two-nonzero target field samples, its shape was still indicative of the proposed azimuth-Doppler characterization of the target, particularly in the contour map of the recovered target field.

While maintaining the rest of the simulation parameters unchanged ($K = 31$, $M = N = 6$), the degradation of the recovered target field in the presence of progressively larger levels of clutter power was observed. The emergence of smaller target field components at locations other than that of the true target was realized. Studying these artifactual target field components, they were found to occur in association with areas of the Azimuth-Doppler Spectrum that presented a structure that resembled that of a valid target. Typically, these areas occurred at points in the Azimuth-Doppler Spectrum where the clutter ridge intersects other spectral elements, such as the side lobes of the target spectral structure, or at the extremes of the ridge, particularly if they were at a corner of the Azimuth-Doppler Spectrum. In the case of target side lobe intersections with the clutter ridge, the local superposition of power generates what resembles a target main lobe, and the elements intersecting at that Azimuth-Doppler combination appear to play the role of the side lobes for the artifactual "main lobe". In the case of the extremes of the clutter ridge, if there is a significant accumulation of power at the corner of the Azimuth-Doppler Spectrum the process of deconvolution seems to resolve the missing information for those power structures (i.e., the missing $\frac{3}{4}$ of the surrounding spectrum) in favor of identifying them as matches with the kernel used for the deconvolution (the spectrum of an elementary target).

It should be stressed that the understanding gained of the possible generation of secondary components in the target field retrieved through deconvolution can be used to develop post-processing mechanisms to sort-out those components that are likely to be artifactual. For example, if a larger target field component is identified, smaller components along that same azimuth and Doppler values could be suspected as being

artifactual. The highly simplified structure of the target field, with respect to the original Azimuth-Doppler Spectrum, would allow for validation mechanisms of this type to be executed rapidly.

The next stage of testing involved working with Azimuth-Doppler spectra at a higher resolution ($K = 48$). This represented a significant increase of the processing time, since it involved a size increase of the convolution matrix H from dimensions 31^2 -by- 31^2 to dimensions 48^2 -by- 48^2 . However, this was necessary so that experimentation with realistic $N=14$ and $M=16$ parameters could be performed. Since the number of lobes in the impulse response or kernel of the transformation in each dimension of the Azimuth-Doppler Spectrum is given by N and M , the higher number of lobes can only be adequately represented if the resolution of the spectrum is also higher. In fact, even at a resolution $K = 48$, the main lobe of the kernel generated assuming $N = 14$, $M = 16$ is represented by a smaller number of Azimuth-Doppler spectral values than for $N=M=6$, at $K=31$. This reduction in effective resolution will make the process of deconvolution more susceptible to mismatch between the spectral expression of a target and the kernel itself. This mismatch seems to result in a target indication in the target field that appear to be more "fragmented", comprising a number of peaks clustered together (as oppose to a single peak). This effect, however, impacted only the NNLS implementation of deconvolution, which also was observed to be more susceptible to the generation of secondary peaks at the intersections of target side lobes and the clutter ridge, as explained above. Van Cittert's method, on the other hand, seemed to be much less affected by the possible mismatch between the actual target and the kernel. This method, in fact, continued to produce thin pulses indicating the target characteristics. Interestingly, the number of non-zero target field values clustered in the resulting thin pulse seems to be about the same as for the lower resolution ($K = 31$) examples. In that sense, then, the indications obtained from Van Cittert's deconvolution were more specific at this higher resolution ($K = 48$).

The effectiveness of a Sample Matrix Inversion (SMI) adaptive spatial filter for removing jammer interference prior to deconvolution was tested at the resolution of $K = 48$, for two different levels of clutter power, 0 dB and 30 dB. In preliminary studies during the preceding Summer Faculty Research Program project, we had detected that the spectral structure of barrage jammer was recovered by the deconvolution process as a line of target indicators along the same jammer azimuth value, causing the process not to retrieve valid target information. We fulfilled one of the proposed goals of this SREP project by incorporating the SMI adaptive spatial filter, with diagonal loading, as a pre-processor for the deconvolution approach. The operation of this pre-processor is not computationally expensive. This operation was outlined in the section on methodology of this report.

The results obtained from a simulated spectrum in which a significant jammer component had been added to the clutter and target components were positive. No artifactual target field peaks were developed at the angle of the jammer when the data coming out of the SMI pre-processor was deconvolved. For the case

including low-power clutter (0 dB) the only elements retrieved in the target field were those associated with the actual target. When a stronger clutter component was simulated (30 dB) the results were similar to the case for those levels of power in the target and the clutter, without a jammer component. It was observed, however, that the SMI pre-processing seemed to accentuate the mismatch between target spectrum and kernel, as slightly more secondary peaks were generated by NNLS deconvolution in the target field. Van Cittert's deconvolution was also affected by SMI pre-processing for the stronger clutter case, generating one spurious component, at the corner of the target field.

The last set of tests carried out were performed with a resolution of $K = 48$, on data retrieved from the Common Research Environment for STAP Technology (CREST) Datalibrary. The files in the datalibrary were recorded under the DoD MountainTop Program. Most of the files of interest were collected with the Radar Surveillance Technology Experimental Radar (RSTER) system, using 14 elements on receive. The RSTER system is really mounted on a stationary platform. So, to emulate the clutter dispersion in the Doppler domain that is characteristic of moving platforms, the concept of Inverse Displaced Phase Center Array (IDPCA) is used. The IDPCA system transmits subsequent pulses from different columns, in rapid succession, thus emulating the transmission of pulses from a single source in a platform that is effectively moving [8]. The specific file for which deconvolution results are shown, "stap 1004v1.1030, cpi 3724, range bin 20". Was recorded with the RSTER and the IDPCA systems mounted at North Oscura Peak, in the White Sands Missile Range. In this particular case there is a mountainous are, called "Diamond Peak", at a distance of about 154 Km from the RSTER, spanning an azimuth interval around -20° (with respect to the broadside of the RSTER at the time this record was obtained). For this test, a moving target simulator (MTS) was set up to introduce a target signal very close to the strong clutter patch, at slightly less negative azimuth and Doppler. In these more realistic conditions the Van Cittert's method of deconvolution seemed to perform best, generating a correct target indication and minimizing the strung clutter scatterer to a smaller target field peak that, in addition, appears now clearly separated from the target indication. In this case the NNLS method of deconvolution was also capable to generate its largest component at the correct target Azimuth and Doppler. However, NNLS generated more secondary peaks, including one that represented a second strong clutter patch at positive Doppler and azimuth.

Conclusions

Through this Summer Research Extension Program project we have been able to further explore the deconvolution of the Azimuth-Doppler Spectrum (ADS) of signals from an airborne multi-element array system, for moving target detection. One of the aspects that makes this approach of particular interest is that it is based on retrieving information about the target from the ADS, as opposed to other techniques, such as Space-Time Adaptive Processing (STAP), which concentrate in the identification of interference components (clutter and jammer) for their cancellation. As such, the deconvolution approach can be used in conjunction with STAP, to complement it or, possibly, to guide the selection of Azimuth-Doppler cells that STAP scrutinizes, by assigning corresponding space-time steering vectors for evaluation.

This study confirmed the feasibility of using a simple pre-processor, such as the Sample Matrix Inversion spatial filter with diagonal loading, to remove jamming components in the ADS before attempting its deconvolution. Our preliminary study (SFRP) had revealed that, while deconvolution is capable of rejecting the clutter interference, it is not robust against significant jammer components.

In dealing with Azimuth-Doppler spectra containing target and clutter components, the deconvolution concept was shown to have the capacity to retrieve the target information. The clutter interference was canceled or attenuated significantly for cases where the power of the background returns was moderately higher than the target to be retrieved (30 to 40 dB stronger).

In this project we explored the deconvolution of data at progressively higher levels of realism. Both deconvolution approaches tested, NNLS and Van Cittert's method, were able to isolate the target information in simulated cases where the relative spectral resolution allowed a good match between the proposed "impulse response" or "kernel" and the ADS expression of the target ($N=M=6$, $K=31$). When more standard values for the number of antenna elements and pulses per CPI were used ($N=14$, $M=16$), an increase in the ADS resolution was needed ($K=48$), at the expense of the associated increase in computation time. In these cases the potential mismatch between the kernel and the target expression in the ADS resulted in the appearance of some secondary components in the recovered target field. This was observed mainly in the intersection of spectral structures due to the target and the clutter, or in the corner of the ADS. The insight gained about these secondary peaks leads us to the conclusion that they can be discarded by a relatively simple post-processor.

Both forms of deconvolution implemented in this project were able to retrieve information about the target in the test with the CREST data shown. The Van Cittert's method seemed to be less affected by possible kernel-target mismatch and also performed better in the CREST data test. This method has an added advantage in that it can be set to perform a pre-determined number of iterations, each one of them involving a known number of operations, which makes its timing predictable and dependent only on the ADS

resolution. The Non-Negative Least Squares (NNLS) method, on the other hand, has a computation time that is data-dependent and varies, even for a fixed ADS resolution level.

Having available the C implementation of both deconvolution methods, as a result from this project, our group is currently contemplating the parallelization of both of them, which could bring their execution times closer to the real-time requirements determined by the application. This will also enable us to study a larger number of cases and variations of the methods, possibly at higher levels of ADS resolution.

References

1. Barile C., Guella T. P. and Lamensdorf D., " Adaptive Antenna Space-Time Processing Techniques to Suppress Platform Scattered Clutter for Airborne Radar", IEEE Transactions on Aerospace and Electronic Systems, Vol. 31, No. 1, January 1995, pp. 382 - 389.
2. Carlson B., "Covariance Matrix Estimation Errors and Diagonal Loading in Adaptive Arrays". IEEE Transactions on Aerospace and Electronic Systems, Vol. 24, No. 4, July 1988, pp. 397-401.
3. Carlson B., Goodman L., Austin J., Ganz M. and Upton L., "An Ultralow-Sidelow Adaptive Array Antenna". The Lincoln Laboratory Journal, Vol. 3, No. 2 (1990).
4. Compton, Jr., "Adaptive Antennas, Concepts and Performance", Prentice-Hall, 1988
5. Eaves J. and Reedy E., "Principles of Modern radar". Van Nostrand Reinhold Company, 1987.
6. Lawson and R. Hanson, " Solving Least Squares Problems", Prentice-Hall Series in Automatic Computation, 1974.
7. Lucy, L.B., "An Iterative Technique for the Rectification of Observed Distributions", Astronomical Journal, Vol.79, No. 6, June 1974.
8. Marshall, D.F., "Preliminary STAP Performance with Mountaintop IDPCA Data", MIT Lincoln Laboratory, Proceedings of the Adaptive Sensor Array Processing (ASAP) Workshop, March 1994.
9. Raemer, H. R., "Radar System Principles", CRC Press, 1997.
10. Richardson, W. H., "Bayesian-Based Iterative Method of Image Restoration" Journal of the Optical Society of America, Vol. 62, No. 1, January 1972.

11. Selitkar Y., Williams D. and McClellan J., "Evaluation of Partially Adaptive STAP Algorithms on the Mountain Top Data Set". ICASSP '96 (pre-print).
12. Skolnik (ed.), "Radar Handbook, 2nd. Edition", McGraw-Hill Publishing Co., 1990.
13. Titi, "An overview of the ARPA/NAVY Mountaintop Program," 1994 IEEE Long Island Section Adaptive Antenna Systems Symposium, Melville, New York, November 7-8, 1994.
14. Titi G. and Marshall D., "The ARPA/NAVY Mountaintop Program: Adaptive Signal Processing for Airborne Early Warning Radar". Proceedings of the IEEE ICASSP '96, May 1996.
15. Van Cittert, P.H., Z. Physik, Vol. 69, p. 298, 1931
16. Wang and L. Cai, "On adaptive Spatio-Temporal Processing for Airborne Surveillance Radar Systems", IEEE Transactions on Aerospace and Electronic Systems, Vol. 30, No. 3, July, 1994, pp. 660 - 669.
17. Ward J., "Space-Time Adaptive Processing for Airborne Radar", Technical Report # 1015, Lincoln Laboratory, MIT, Lexington Massachusetts, December 1994.

APPENDIX: C-implementation for Van Cittert's Deconvolution

(NOTE: The length of the C-code for NNLS deconvolution makes it impractical to include it in this appendix. The code, however, is available from the principal investigator at: barreto@eng.fiu.edu

```
/******Van_Cittert's_Deconvolution_of_a_48-by-48 Azimuth-Doppler Spectrum******/

#include <stdio.h>
#include <math.h>
#include <stddef.h>
#include <time.h>
#include <cblas.h>

#define MAXSTRING 55
#define MAX 2304
#define SQ_MAX 48
#define MAX_SQ 5308416
#define MAX_ITER 100

/****** Global Variable Definitions ***** */
double values[MAX];
/* This array will hold the
/* column vector values read
/* in from the text files created
/* in Matlab (target and PSD)

double bvalues[MAX_SQ]; /* This array will hold the big
/* column vector values read in from
/* the text files created in Matlab

double bmatrix[MAX][MAX]; /* This array will hold the
/* square matrix values
/* reshaped

int length[2]; /* lengths of column vectors for:
/* 0 = power spectral density (y),
/* 1 = point spread function (H)

void main(int argc, char **argv)
{
/****** Variable Definitions ***** */
int m, n, o, i, j, k, iteration, count, N, M, incx, incy;

double target_0[MAX], target_1[MAX], target_m[MAX], residue[MAX],
cir_conv[MAX], radius;
double alpha, local_temp, temp, acc, total_time, dN, c1, c2;
time_t timer, t1, t2;

FILE *outdata, *indata[2];

/****** Check if correct number of arguments were passed ***** */
if (argc < 5)
{
fprintf(stderr, "Usage:\t%s\tPSD file\tPSF (H)
file\tRecovered Target alpha\n", argv[0]);
exit(0);
}
alpha = atof(argv[4]);
```

```

/***** Read values contained in 1st two argument files *****/
for (n=0; n<2; n++)
{
    indata[n] = fopen(argv[n+1], "rb");
    length[n] = 0;
    count = length[n];

    if (n==0)
        while (fscanf(indata[n], "%lf\n", &values[count]) == 1)
        {
            length[n]++;
            count = length[n];
        }
    else
        while (fscanf(indata[n], "%lf\n", &bvalues[count]) == 1)
        {
            length[n]++;
            count = length[n];
        }

    dN = sqrt(length[n]);
    N = (int)dN;
    printf("\nNumber of values in %s = %d\n", argv[n+1],
        length[n]);
    printf ("Square Root of %d = %d\n", length[n], N);
}

/***** Reshape column vector into a square matrix for dot product*****/
for (i=0; i<N; i++)
    for (j=0; j<N; j++)
    {
        k = (N*i) + j;
        bmatrix[i][j] = bvalues[k];
    }

/***** Initiallize target matrix to PSD read in from file*****/
for (i=0; i<length[0]; i++)
    target_0[i] = values[i];

n = length[0];
incx = 1;
incy = 1;

/***** Start time for Van Cittert's Algorithm *****/
t1 = time(&timer);

for (iteration=0; iteration < MAX_ITER; iteration++)
{
    for (i=0; i<length[0]; i++)
    {
        cir_conv[i] = ddot(n, bmatrix[i], incx, target_0, incy);
        residue[i] = values[i] - cir_conv[i];
    }
    for (i=0; i<length[0]; i++)
    {
        target_1[i] = target_0[i] + (alpha * residue[i]);
        target_0[i] = (target_1[i] + fabs(target_1[i])) / 2;
    }
}

/***** Stop time for Van Cittert's Algorithm *****/
t2 = time(&timer);

/***** Pass calculated data to final argument *****/
outdata = fopen(argv[3], "wb");
for (i=0; i<length[0]; i++)
    fprintf(outdata, "%f\n", target_0[i]);

printf("elapsed time %d seconds\n", t2-t1);
}

```

Modeling and Design of New cold Cathode Emitters
Using Wide Bandgap Semiconductors

M. Cahay
Associate Professor
Department of Electrical Engineering

University of Cincinnati
Cincinnati, OH 45221

Final Report for:
Summer Research Extension Program
Wright Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

and

University of Cincinnati

December 1997

Modeling and Design of New Cold Cathode Emitters Using Wide Bandgap Semiconductors

M.Cahay
Associate Professor
Department of Electrical Engineering
University of Cincinnati

ABSTRACT

We analyze the importance of current crowding in a new cold cathode emitter which consists of a thin wide bandgap semiconductor material sandwiched between a metallic or heavily doped semiconductor and a low work function semimetallic thin film. Potential material candidates are suggested to achieve low-voltage (< 10 V), room-temperature cold cathode operation with emission currents of several tens of A/cm^2 . We calculate the lateral potential drop which occurs across the emission window of cold cathodes with circular geometry and describe its effects on the emitted current density profile. The power dissipation in the cold cathode is calculated as a function of a dimensionless parameter characterizing the importance of current crowding. We determine the range of dc bias over which cold cathodes of different radii must be operated to minimize current crowding and self-heating effects.

I. INTRODUCTION

Recently, there has been renewed interest into cold cathode emitters for applications to a variety of electronic devices, including microwave vacuum transistors and tubes, pressure sensors, thin panel displays, high temperature and radiation tolerant sensors, among others [1, 2]. Introduction of such emitters would permit an unprecedented compactness and weight reduction in device and equipment design. Low temperature operation in nonthermionic electron emitters is very desirable for keeping the statistical energy distribution of emitted electrons as narrow as possible, to minimize thermal drift of solid state device characteristics, and to avoid accelerated thermal aging or destruction by internal mechanical stress and fatigue. To keep the emitter temperature rise small appears easy if the emitters are built as thin epitaxial films using vertical layering technology due to the extremely short heatpaths and excellent heatsinking possibilities offered with this architecture. For an electron emitter to be useful in microwave tube applications it should be capable of delivering current densities in excess of 10 A/cm^2 and to sustain emission during operational lifetimes over periods of 10^5 hrs. To satisfy this requirement, the structural and chemical composition must be stable. This rules out the historically practiced use of alkali metal films on emitter surfaces for the lowering of electronic work functions. These films sublime, evaporate or surface migrate over time and end up on various surfaces inside the vacuum envelop.

Several cold cathode emitters have been proposed since their first successful demonstration by Williams and Simon [3] using a cesiated p-type GaP structure. A review and criticism of the different cold cathode approaches was given recently by Akinwande et al. [4]. In this work, we propose a new cold cathode emitter concept and use a simple model to show that the new emitter is capable of achieving low voltage ($< 10 \text{ V}$) room temperature operation with emission current approaching 100 A/cm^2 and large efficiencies. A preliminary report of this work has been published earlier [5]. The architecture of the structure is shown in Fig. 1. The main elements in the design and functioning of such an emitter are : (1) a wide bandgap semiconductor slab equipped on one side with a metallic contact [6] or a heavily doped semiconductor ($n^{++} - \text{InP}$) on one side of an undoped CdS region that supplies electrons at a sufficient rate into the conduction band and (2) on the opposite side, a thin semimetallic film that facilitates the coherent transport (tunneling) of electrons from the semiconductor conduction band into vacuum. Of importance is the mutual alignment of the crystalline energy levels at the semiconductor-semimetal film junction. This requires the use of new materials and development of their epitaxial growth technologies. For that

reason, the choice of InP as a substrate is particularly attractive since the lattice constant of InP (5.86 Å) closely matches the lattice constant of the zincblende cubic CdS (5.83 Å). Furthermore, there have been recent reports on the deposition of crystalline layers of CdS on InP by molecular beam epitaxy [7], chemical bath deposition [8], and pulsed laser deposition [9]. The proposed cold cathode should therefore be realizable with present day technology.

As shown in Fig. 1(a), a thick metal grid is defined on the surface of the LaS thin film to bias the structure. There are openings in the grid structure to expose the thin LaS film which forms the active emission area of the cold cathode. Cathodes with rectangular (Fig.1(b)) emission windows were studied previously [5]. Current crowding and self-heating effects in cathodes with circular geometry (Fig.1(c)) emission windows will be considered hereafter. The bias is applied between the back metallic contact and the metal grid with emission occurring from the exposed LaS surface. If the applied voltage is equal or larger than the semiconductor bandgap energy and the quotient of the applied voltage divided by the semiconductor thickness approaches $0.1\text{eV}/\text{\AA}$, then electrons are tunnel injected into the conduction band and ascend during their travel across the semiconductor film to levels of increasing energy. Referring to Fig. 2, the conduction band of the wide bandgap semiconductor provides the launching site for electrons where they are - through a thin film - injected into vacuum. This injection of electrons into vacuum becomes possible and is effective as long as the semimetallic film is very thin and has a work function small enough so that its vacuum edge is located energetically below the conduction band edge of the semiconductor. This situation is referred to as negative electron affinity (NEA) for the semiconductor material [10]. Depending on the particular materials choices, this implies that the semimetal work function ϕ_M in relation to the semiconductor energy bandgap E_G must obey one of the inequalities $\phi_M < 0.5E_G$ or $\phi_M < E_G$ if an intrinsic or p-type doped wide bandgap semiconductor is used, respectively. A negative ϕ_M implies according to Fig.2 that the vacuum level would be located below the lower conduction band edge. In that case, electrons in the conduction band with momenta pointing toward the surface have a good chance to get emitted unless deflected by collision or trapped by impurities or defects.

This paper is organized as follows. In section II, we derive the basic equations describing the forward bias operation of the cold cathode emitter described above. We then calculate the current density-voltage characteristics of the newly proposed cold cathode for specific sets of materials and device parameters. In section III, we investigate the importance of current crowding effects in various cold cathodes with circular geometry. Our analysis includes a

self-consistent modeling of current crowding effects and an analysis of power dissipation in the cold cathode active area. The influence of power dissipation on self-heating effects in the active area of the cathode is also described. Finally, Sec. IV contains our conclusions.

II. THE MODEL

Hereafter, we analyze the cold cathode whose energy band diagram is shown in Fig. 2 [5]. Under the influence of a large electric field in the wide bandgap semiconductor, electrons will eventually tunnel from the left contact through the barrier at the metal-semiconductor interface. A portion of the current emitted at the metal or heavily doped semiconductor -*CdS* contact (which we model assuming Fowler-Nordheim injection) is transmitted at the boundary of the *LaS* as well as the vacuum boundary. However, a fraction of the current is lost in the thin *LaS* quantum well gives rise to the dynamic shift of the effective material work function (Fig. 2). For a cathode operated at room temperature, we model this internal field emission at the injection junction using a Fowler-Nordheim (FN) type expression for the injected current (in A/cm^2) [11]

$$J_{FN} = C_1(E^2/\Delta)e^{C_2\Delta^{3/2}/E}, \quad (1)$$

where C_1 and C_2 are constants which depend on the wide bandgap semiconductor. In our numerical simulations, we chose $C_1 = 1.5 \times 10^6$ A/V and $C_2 = 6.9 \times 10^7 (V^{1/2}cm)^{-1}$ which are of the same order of magnitude as the constants appearing in the FN expression [11]. In Eq.(1), Δ is the barrier height (in eV) at the metal-semiconductor junction and E is the electric field (in V/cm) in the wide bandgap semiconductor [12]. We assume that the semiconductor layer thickness is such that the transport of injected electrons is close to being ballistic up to the interface between the semiconductor and the thin semimetallic film. In so doing, we also neglect carrier ionization processes in the semiconductor slab which could be the main antagonist to ballistic transport in that region.

Because of the finite probabilities for the injected current to be transmitted at the semiconductor-semimetal (probability T_1) and semimetal-vacuum interfaces (probability T_2), the contributions to the total emitted current can be calculated as the sum of the contributions resulting from the multiple reflections of electrons in the semimetallic layer (See Fig. 2). The magnitudes of the emitted current components decreases with the number of multiple reflections in the semimetallic layer. Rather than trying to calculate these contributions exactly, we assume that the current amplitude is decreased by a factor $\epsilon = \exp(-L_2/\lambda_{LaS})$

for each traversal of the semimetallic layer, where λ_{LaS} is the collisional mean free path in the semimetallic layer and L_2 is the length of the semimetallic layer. Adding the contributions resulting from multiple crossings of the semimetallic layers, the total emitted current is found to be

$$J_{em} = \epsilon T_1 T_2 J_0 (1 + x + x^2 + \dots), \quad (2)$$

where $x = \epsilon^2 (1 - T_1)(1 - T_2)$. In calculating J_{em} we limited the number of traversals of the semimetallic slab to five to include the fact that electrons loose energy in each crossing and eventually do not have enough energy to surmount the barrier at the semimetal-vacuum interface. According to Eq.(2), the contributions from the multiple reflections decrease rapidly since, in general, the quantity x will be much smaller than unity [13]. Once the emitted current is found, the total current contributing to the increase in the sheet carrier concentration in the thin semimetallic film can easily be written as $J_{capt} = J_{FN} - J_{em}$. The total trapped current is then given by

$$i_T = A J_{capt} = A R J_{FN} \quad (3)$$

where A is the area of each LaS emission window which in practice can be either rectangular or circular (See Fig. 1b and 1c). In Eq.(3), R is the trapping coefficient of the well

$$R = 1 - \epsilon T_1 T_2 (1 + x + x^2). \quad (4)$$

The semimetallic thin film can be modeled as a quantum well (Q.W) which will loose the trapped electrons essentially at its lateral boundaries. In reality, Fig. 1(a) indicates that not all electrons will move to the three-dimensional contact regions surrounding the thin semimetallic layer but many of them will get reflected at the lateral thin film layer with an average probability r (calculated for electrons with the Fermi velocity in the thin film). The exiting number of electrons will depend on the thickness of the semimetallic layer and could be adjusted by intentional passivation so that reflection at the boundaries of the thin semimetallic film could be tuned from almost zero to nearly unity. Taking into account the finite reflection amplitude at the thin film boundaries, the leakage current of the Q.W can be rewritten

$$\frac{dQ_T}{dt} = 2e L N_2 D v_F (1 - r), \quad (5)$$

for the case of a rectangular emission window and

$$\frac{dQ_T}{dt} = 2\pi e a N_2 D v_F (1 - r), \quad (6)$$

for the case of a circular emission window.

In Eqns.(5) and (6), Q_T is the total charge captured by the well, e is the magnitude of the electronic charge, N_{2D} is the excess sheet carrier concentration in the thin film due to the captured electrons, and v_F is the Fermi electron velocity in the semimetallic thin film. Under steady state operation of the cold cathode, the excess charge in the two-dimensional semimetallic film is found using Eq.(3) and imposing the current balance requirement $\frac{dQ_T}{dt} = i_T = AJ_{capt}$. This leads to

$$N_{2D} = WJ_{capt}/2e(1 - r)v_F, \quad (7)$$

for the case of a rectangular geometry and

$$N_{2D} = aJ_{capt}/2e(1 - r)v_F, \quad (8)$$

for the case of a circular geometry.

Simultaneously, the change N_{2D} in the excess sheet carrier concentration in the Q.W due to trapped electrons leads to the occupation of the boundstate energy levels according to the energy density of states up to an energy level which will establish the dynamic Fermi level E_F^1 . The Fermi velocity v_F entering Eqns.(5) and (6) must be calculated self-consistently because of the dynamic work function shift $|\Delta\chi|$ illustrated in Fig. 2. This dynamical shift $|\Delta\chi|$ is equal to $|E_F^1 - E_F^0|$, where E_F^0 is the Fermi level in the thin semimetallic layer under zero bias. For simplicity, we assume that the electrons in the conduction band of the semimetallic films can be described using the Sommerfeld theory of metals while assuming s-band conduction in the semimetallic thin film and while modeling the thin film using the particle in a box model for the quantum well [14]. The set of equations (1-8) is then solved self-consistently to calculate the work function shift $|\Delta\chi|$ as a junction of the externally applied bias. Once the dynamic shift has been determined self-consistently, Eq.(2) can then be used to determine the emitted current.

RESULTS

We consider a specific structure with the material and structural parameters listed in Table I and II, respectively. Both Au and Ag are known to form contacts to thin films of semiconducting (n-type) CdS. In that case, the barrier height Δ shown in Fig. 2 is quite small and is equal to 0.78 eV and 0.56 eV for the case of Au and Ag contacts, respectively [11]. The lattice constant of CdS (5.83 Å) is very close to the lattice constant of the thin semimetallic surface layer LaS (5.85 Å) which in its cubic crystalline structure will therefore

be lattice matched to the semiconducting material. Additionally, LaS is expected to have quite a low room temperature work function (1.14 eV) [15], a feature when combined with the large energy gap (2.5 eV) of CdS leads to NEA of the semiconductor material. In the following numerical simulations, the thicknesses of the CdS (L_1) and LaS (L_2) layers are set equal to 500 Å and 24.6 Å (4 monolayers), respectively. We model a cathode with a square ($W = L$) emission window with a 1 cm^2 area.

Figure 3 is a plot of the dynamic work function shift as a function of applied bias for the cold cathode emitter with both Au and Ag injecting contacts. The following parameters were used: $\lambda_{LaS} = 300 \text{ Å}$, $T_1 = T_2 = 0.5$, and $v_F = 1.36 \times 10^8 \text{ cm/s}$. Figure 3 indicates that the dynamic shift of the LaS work function is sensitive to the quality of the interface between the two-dimensional semimetallic layer and the three-dimensional contacts which we model by varying the reflection coefficient r between the two-dimensional semimetallic thin film and the three-dimensional contact regions (See Fig. 1(a)). It should be noticed that the LaS work function shift can approach the LaS workfunction even for the case of a *leaky* interface between the thin semimetallic layer and the 3D contact regions. The dynamic shift $|\Delta\chi|$ is comparable to the work function of LaS for a smaller value of the applied bias in the case of Ag contact because of the lower barrier at the Ag/CdS interface.

Figure 4 compares the emitted current densities J_{em} for the structure with Au and Ag contacts calculated while including or neglecting the effects of the dynamic shift of the LaS work function. The current density versus bias plots are stopped at the values of V_{bias} at which $|\Delta\chi| = \phi_M(LaS) = 1.14 \text{ eV}$. Beyond that point, the theory exposed here is no longer valid since we would need to include the spill over of the excess trapped carriers into vacuum. As can be seen in Fig. 4, the emitted current densities can be more than a factor two larger when the effects of the dynamic shift of the work function of the semimetal are included. The effects could be made more drastic if a set of materials and device parameters could be found for which the dynamic shift of the work function could be made comparable to the work function itself at fairly low value of the applied bias ($< 5V$).

Sensitivity of Dynamic Work Function Shift on Design Parameters

The previous numerical examples have shown that, under forward bias operation, the electrons captured in the low work function material are responsible for an effective reduction of the semimetallic film work function together with a substantial increase of the cathode emitted current. This dynamic work function shift was shown to increase with the amount of

injected current. Hereafter, we perform a more extensive study of the dynamic work function shift which includes variations of the length of the *CdS* region (L_1), the electron mean free path in the *LaS* region (λ_{LaS}), the emission window size (W), the transmission coefficients at the *CdS/LaS* (T_1) and *LaS/Vacuum* interfaces (T_2), and the reflection at the 2D/3D interface region in the *LaS* quantum well (r).

Figure 3 indicates that the dynamic work function shift $|\Delta\chi|$ rises exponentially above a threshold voltage of several volts and reaches rapidly (within a few volts range) a value comparable to the *LaS* work function. For a structure with the parameters listed in Table I and with the structural and physical parameters ($L_1 = 500 \text{ \AA}$, $L_2 = 24.6 \text{ \AA}$, $W = 1 \text{ cm}$, $\lambda_{LaS} = 300 \text{ \AA}$, $T_1 = T_2 = 0.5$, $\Delta(A_g) = 0.56 \text{ eV}$), Fig. 5(a) shows that the dynamic work function shift rises exponentially at a lower bias as the reflection coefficient at the 2D/3D interface in the *LaS* region approaches unity. Figure 5(b) also shows that the difference between the current densities calculated with and without including the dynamic work function shift are more pronounced for smaller values of the applied bias when the reflection coefficient between the 2D and 3D *LaS* regions is approaching unity. This results from the fact that any mechanism (like r being closer to unity) which increases the amount of charge being trapped in the *LaS* quantum well leads to an enhancement of the dynamic work function shift at a given bias. For instance, all other cathode parameters being equal, the dynamic work function rises much faster as a function of applied bias in structures with thinner *CdS* regions (Fig. 6), with smaller values of the transmission coefficients T_1 and T_2 (Fig. 7), or with smaller mean free path (λ_{LaS}) in the semimetallic thin film (Fig. 8). As shown in Fig. 8(a), the dynamic work function shift occurs at a lower bias as λ_{LaS} is decreased. On the other hand, the emitted current density is lesser at a given bias in a cathode whose *LaS* thin film has a lower electron mean free path (Fig. 8(b)). We have found this trend to be valid for all values of the reflection coefficient at the *LaS* 2D/3D interface. However, there is a larger spread in the family of curves representing the bias dependence of the dynamic work function and emitted current densities as a function of the mean free path in the *LaS* thin film for smaller values of the reflection coefficient r . Finally, even though not shown here, the exponential rise of the dynamic work function shift and emitted current density has been shown to occur at lower value of the bias by either reducing the thickness of the *CdS* layer or by lowering the barrier height Δ at the metal/*CdS* interface.

Before leaving this section, we make two additional remarks on the bias dependence of the emitted current density and the dynamic work function shift which we have checked

numerically on all the cold cathodes modelled in this work. First, since the emission of electrons at the metal(or heavily semiconductor)/*CdS* interface is assumed to be of the Fowler-Nordheim type, the emission current is expected to have a bias dependence of the form

$$J_{FN} = AV_{bias}^2 \exp(-B/V_{bias}). \quad (9)$$

This was checked numerically for all the cold cathodes modelled here. A typical example is given in Fig.9(a), with the values of the parameters *A* and *B* listed in the inset. Most of the quantities of interest to be determined hereafter (power dissipation, lateral variation of emitted current and lateral potential drop,...) can be calculated exactly analytically if the following approximation is used

$$J_{FN} = J_0 e^{\alpha V_{bias}}. \quad (10)$$

For all the cold cathodes simulated here, we have shown that Eq.(10) gives a fairly accurate fit to the plot of the emitted current density versus applied bias if the range of the fit is restricted to current densities between 1 and 1000 A/cm². A typical fit for the one of the cold cathode studied here is shown in Fig.9(b) with the values of the parameters *J*₀ and *α* in Eq.(8) shown in the inset. Table III gives a summary of the parameters *J*₀ and *α* for cold cathodes of different width and with the physical parameters listed in the caption of Fig.5. We have found that the bias dependence of the dynamic work function shift is also of the Fowler-Nordheim type, i.e.,

$$\Delta\chi = \Delta_0 V_{bias}^2 \exp(-V_0/V_{bias}). \quad (11)$$

This is illustrated in Fig.10 for the cold cathode with the same parameters as in Fig.9. The values of *Δ*₀ and *V*₀ are indicated in the inset of Fig.10. We point out that the values of the parameters *B* and *V*₀ in Eqns.(9) and (10) are nearly identical.

Temperature Rise in the Cold Cathode

Hereafter, we derive an upper estimate of the temperature rise in the *LaS* thin film as a result of the power dissipation mechanisms discussed above. We focuss on a cold cathode where the *CdS* thin film is deposited on a *InP* substrate as shown in Fig.1(a). The successful growth of cubic *CdS* thin films with good crystalline quality on *InP* substrates has been reported recently by Shen and Kwok [9]. For the case of a *InP*/*CdS* interface, there has not been any report of the conduction band discontinuity *ΔE_c* at the interface between the two

materials, to the best of our knowledge. For that reason, ΔE_c was assumed to be given by Anderson's rule, i.e, Δ in Fig. 2 is assumed to be given by $|\chi(InP) - \chi(CdS)| = 0.2$ eV [16]. This estimate was based on the measured electron affinities of *InP* (4.4 eV) and *CdS* (4.2 eV) reported in refs. [17] and [18], respectively. The back contact to the substrate is assumed to be perfectly ohmic and to act as a perfect heat sink (300K). Since the device area (heat source formed of the *LaS* thin film) is much thinner than the substrate, it is necessary to consider the effect of heat spreading laterally in the substrate. Furthermore, we assume that the temperature of the *CdS* layer will be the same as the *LaS* top layer. The active area of the cathode (*CdS* and *LaS* layers) is therefore assumed to be acting as a heat source with the power density calculated in the previous section. Because of the finite thermal conductivity of the substrate, we expect self-heating effects to affect the operation of the cold cathode if the power level dissipated in the active area of the cathode becomes too important. Hereafter, we model the thermal conductivity of the *InP* substrate as follows

$$\kappa(T) = \kappa_0(T/T_0)^{-b}, \quad (12)$$

where κ_0 is the thermal conductivity at T_0 (300K), $\kappa_0 = 0.74$ W/Kcm is the room temperature thermal conductivity of *InP* and $b=1.45$. Starting with Fick's law and making use of a Kirchhoff transformation to take into account the temperature dependence of the thermal conductivity of the *InP* substrate given by Eq.(12), it can be shown that the active area of the cold cathode will be operated at a temperature given by

$$T = \left[\frac{1}{T_0^{b-1}} - (b-1) \frac{R_{th,0} P_{diss}}{T_0^b} \right]^{\frac{-1}{(b-1)}}, \quad (13)$$

where T_0 is the ambient room temperature (assumed to be 300 K hereafter), P_{diss} is the total power dissipated per finger as calculated in the previous section, and

$$R_{th,0} = \frac{1}{\kappa_0} \int_0^{z_s} \frac{dz}{A(z)}, \quad (14)$$

where Z_s is the thickness of the *InP* substrate. Our estimate of the temperature rise in the active cold cathode area will give an upper estimate of the temperature of operation since we neglected heat conduction to the top *Au* contacts in the thick portion of the *LaS* thin film.

III. Current Crowding Effects in Proposed Cold Cathode with a Circular Geometry

In this section, we study the effects of current crowding in the case of an emission window with circular geometry (see Fig.1(c)). The lateral potential drop in the circular LaS window satisfies the following differential equation [19]

$$\frac{dV}{dr} = \frac{\rho_s}{t} i(r), \quad (15)$$

where $i(r)$ is the total lateral current per unit length flowing outward across a circle of radius r , whose center coincide with the center of the emission window. The lateral current satisfies the following equation

$$2\pi r i(r) = 2\pi R \int_0^r dr' r' j(r'). \quad (16)$$

If we further assume that the Fowler-Nordheim emitted current $j(r)$ can be approximated by Eq.(10) over the range of dc bias considered here, the following second-order differential equation must be satisfied by the lateral potential drop:

$$\frac{d^2 V}{dr^2} + \frac{1}{r} \frac{dV}{dr} = \frac{\rho_s R J_0}{t} e^{V(r)/V_T}. \quad (17)$$

This differential equation must be solved subject to the following boundary conditions valid for the circular geometry

$$\frac{dV}{dr} = 0, \quad (18)$$

at $r = 0$, and

$$V(r = a) = V_{bias}, \quad (19)$$

at the edge of the circular window. Introducing the reduced variable $r' = r/a$ and the quantity

$$Y = \frac{V - V_{bias}}{V_T}, \quad (20)$$

Eq.(17) can be recast as follows

$$\frac{d^2 Y}{dr'^2} + \frac{1}{r'} \frac{dY}{dr'} = 2\gamma^2 e^Y, \quad (21)$$

where $\gamma^2 = \frac{\beta^2 a^2}{2}$ and the parameter β is identical to the one defined for the planar problem [20].

The general solution of Eq.(21) can be found analytically and is given by

$$e^Y = (c\delta^2/\gamma^2)[r'^{\delta-2}/(1 - cr'^{\delta})^2], \quad (22)$$

where δ and c are constants to be determined so the boundary conditions (18) and (19) are satisfied. Using the new system of variables, Eq.(18) becomes

$$\frac{dY}{dr'}|_0 = 0, \quad (23)$$

while Eq.(19) now reads

$$Y(1) = 0. \quad (24)$$

Equation (23) leads to

$$\delta = 2. \quad (25)$$

and Equation (24) becomes

$$\delta^2 = (1 - c)^2 \gamma^2 / c. \quad (26)$$

Combining these last two equations, we obtain the following result

$$\gamma^2 = \frac{4c}{[1 - c]^2}, \quad (27)$$

Equation (27) can be solved exactly for the parameter c

$$c = [\gamma^2 + 2 - 2\sqrt{\gamma^2 + 1}]/\gamma^2 \quad (28)$$

in terms of which we can write various quantities of interest, including the ratio

$$J(0)/J(a) = [1 - c]^2, \quad (29)$$

characterizing the importance of current crowding in the circular geometry. Using Eqn.(22), the radial dependence of the lateral potential drop is found to be

$$V(r) = V_{bias} - 2V_T \ln\left[\frac{a^2 - cr^2}{a^2(1 - c)}\right], \quad (30)$$

from which the maximum value of the in plane electric field is found to be

$$E_r(r = a) = \frac{4c}{c - 1} \frac{V_T}{a}. \quad (31)$$

As in the case of the rectangular geometry, the total power dissipated in the *LaS* thin film is given by the sum of the following four contributions [20]. The power dissipated by the electrons being trapped in the *LaS* circular thin film is given by

$$P_1 = 2\pi R \int_0^a dr j(r) V(r). \quad (32)$$

The power dissipated by Joule heating as trapped electrons move to the edge of the *LaS* window is given by

$$P_2 = \frac{2\pi\rho_s R^2}{t} \int_0^a \frac{dr}{r} \left[\int_0^r r' j(r') dr' \right]^2. \quad (33)$$

The third contribution to power dissipation comes from Joule heating linked to the current making it from the *LaS* thin film to the *Au* contacts on top of the thick *LaS* regions.

$$P_3 = R_c (2\pi a i(a))^2. \quad (34)$$

where R_c is the resistance of the *LaS* region between the edge of the *LaS* thin film and the top *Au* layer. This resistance can be estimated as follows [21]

$$R_c = \frac{\rho_s H}{t 2\pi a} \frac{\ln k}{k-1}, \quad (35)$$

where $k = b/t$, a is the radius of the circular window, and H is the height of the thick *LaS* region.

Finally, there is also a contribution to power dissipation due to the blocking effect on the Folwer-Nordheim emission current emitted under the wide *LaS* contacts:

$$P_4 = \pi b(b+2a) J_0 e^{\alpha V_{bias}}. \quad (36)$$

Starting with Eqns.(10) and (30), the different contributions to the power dissipation can be calculated exactly and are found to be

$$P_1 = \pi a^2 J_0 R e^{\alpha V_{bias}} V_{bias} (1-c) - \pi a^2 J_0 R e^{\alpha V_{bias}} V_T \left[2(1-c) + \frac{(1-c)^2}{c} \ln(1-c)^2 \right], \quad (37)$$

$$P_2 = R(\pi a^2 J_0 e^{\alpha V_{bias}}) V_T \left[2(1-c) + \frac{(1-c)^2}{c} \ln(1-c)^2 \right], \quad (38)$$

P_3 is given by Eq.(34) and P_4 is found to be

$$P_4 = \frac{R_c R^2}{2(1-R)^2} I_{em}^2, \quad (39)$$

where I_{em} is the total emitted current through the circular window

$$I_{em} = 2\pi(1-R) \int_0^a r j(r) dr. \quad (40)$$

The latter can be calculated explicitly and is found to be

$$I_{em} = (1-R) [\pi a^2 J_0 e^{\alpha V_{bias}}] (1-c). \quad (41)$$

The input power (per emission window) delivered by the power supply biasing the cold cathode is given by

$$P_{input} = \int_0^{a+b} 2\pi r j(r) V(r) dr, \quad (42)$$

which can readily be shown to be given by

$$P_{input} = P_1/R + P_4. \quad (43)$$

The power efficiency of the cold cathode can be calculated as follows

$$\eta_P = \frac{P_{input} - P_{diss}}{P_{input}}. \quad (44)$$

and the temperature rise in the cathode is given by Eq.(13). where the thermal resistance $R_{th,0}$ must be calculated for the case of power dissipation through the substrate from a heat source with circular geometry. In this case, we find

$$R_{th,0} = \frac{1}{\kappa_0} \frac{z_s}{a} \frac{1}{[a + z_s \tan \theta]}, \quad (45)$$

in which the heat spreading angle θ is set equal to 45° in the numerical examples below, for simplicity.

Numerical Examples

Figure 11 shows the variation of the parameter c in Eq.(28) as a function of applied bias for cold cathodes with circular emission window of different radii. For all cathodes, the physical parameters are the same as listed in the caption of Fig.5. As in the case of a rectangular window, we use the criterion that current crowding is negligible if the lateral potential drop between the center and the edge of the LaS circular window, $V(r=0) - V_{bias}$, is kept less than $0.1V_T$. Using Eqns.(28) and (30), we find that this criterion requires the parameter c to be less than 0.05. This limit is indicated as a vertical line in Fig. 11. The family of curves in Fig.11 is parametrized with the radius of the emission window. Figure 11 shows that the range of dc bias over which current crowding can be neglected in a circular window is comparable to the range of dc bias over which current crowding is negligible in a rectangular window whose width is equal to the radius of the circular emission window [20].

Figure 12 illustrates the importance of current crowding on the lateral potential drop in emitter windows of different radii. The left frames show the radial dependence of the electrostatic potential for four different values (1,10,100,1000 A/cm²) of the current density at

the rim of the circular *LaS* window. The right frames in Fig.12 show the corresponding radial dependence of the emitted current density. From Fig.12, it can be seen that current crowding is negligible in emitter windows with radius less than $50\ \mu\text{m}$ if the emitted current density is kept under $10\ \text{A}/\text{cm}^2$. As in the case of *LaS* windows with rectangular geometry, Figure 12 shows that the current density profiles are much more sensitive to the finite resistivity of the *LaS* thin film than the lateral potential drop.

Figure 13 is a plot of the four contributions to the total power dissipated in cold cathodes with different radii plotted as a function of the parameter c . For all cathodes, the power dissipation due to Joule heating in the *LaS* thin film and the thick *LaS* regions is negligible compare to the power released by electrons being trapped in the *LaS* thin film and by electrons blocked in the thick *LaS* regions. The latter is always about one of magnitude higher than the former. Figure 13 shows that substantial power dissipation occurs in the cathode with radius under $50\ \mu\text{m}$ while the cathode is still operating without any substantial current crowding effects (i.e, $c < 0.05$). The power efficiency η_P of cathodes of different width is plotted as a function of V_{bias} and the emitted current density J_{em} in Figs. 14(a) and 14(b), respectively. For all window size, the efficiency decreases with V_{bias} and J_{em} as a result of current crowding. The efficiency is more or less constant over a wider range of V_{bias} for window with smaller radius because current crowding is less important in that case, as illustrated in Fig. 12. The overall lower efficiency for window with smaller radius illustrated in Fig. 14 comes from the fact that the width of the thick *LaS* regions was set equal to $100\ \mu\text{m}$ for all cathodes. The efficiency of cathodes could be increased by making the ratio b/a in Fig.1(b) closer to unity.

Figure 15 shows the temperature of the active area of a cold cathode with the parameters listed in Table I as a function of V_{bias} for emitter window with different radii. The thickness of the *LaS* contacts and *InP* substrate was set equal to 100\AA and $100\ \mu\text{m}$, respectively. As in case of cold cathodes whose emission window as a rectangular geometry [20], Fig. 15 indicates that to limit the temperature rise in any cathode to less than 200 K, the dc bias must be limited to a smaller range for emitters with smaller window radius. For instance, according to Figures 11 and 15, a cathode with a $20\ \mu\text{m}$ diameter can be operated up to 8.3 V with negligible self-heating effects (ΔT around 100 K). For that bias, Figure 12 indicates that current crowding would be negligible in the cathode and the emitted current density would be around $100\ \text{A}/\text{cm}^2$ (See Fig. 16). On the other hand, Figure 11 indicates that a $100\ \mu\text{m}$ diameter window can be operated up to 6.8 V before current crowding becomes

non negligible. At this bias, the emitted current density would be around 15 A/cm^2 (Fig. 16) while the temperature rise in the device would be only about 15 K as shown in Fig. 15.

IV. CONCLUSIONS

We have proposed a new cold cathode emitter which consists of a thin wide bandgap semiconductor material sandwiched between a metallic material or heavily doped semiconductor, and a low work function semimetallic thin film. We have shown that the capture of electrons by thin semimetallic layers grown on the escape surface of wide bandgap semiconductors can lead to a dynamical shift of the work function of the semimetallic layers together with an increase of the cathode emission current. While varying the device and physical parameters of the structure, our studies suggest that any mechanism which promotes additional charge deposit in the well enhances the dynamic work function shift phenomenon thereby increasing the emitted current. Potential material candidates were proposed for cold cathode operation with applied bias under 10 V, with current densities approaching several tens of A/cm^2 , and with large power efficiencies (η_P approaching 15 %).

The results of our analysis show that a cold cathode with either a rectangular or circular emission window and with the parameters listed in Tables I and II would emit a uniform current density of about 15 A/cm^2 at a dc biasing voltage of about 8V. For that bias, the effects of current crowding would be negligible and the temperature rise in the active area of the cathode (CdS/LaS layers) as a result of self-heating effects would be negligible.

Further improvements to the theory should include a more realistic model for the emission current and transport through the wide bandgap material. Also, the semimetallic film energy density of states (to account for the d-band character of the conduction band in the chosen rare-earth semimetallic samples [14]), the finite probability for electron wavefunctions in the thin semimetallic films to extend in the semiconductor material [22], a more accurate description of the energy loss mechanisms [23] and screening effects (including the lateral ohmic voltage drop) in thin semimetallic layers [24]. Finally, our study of self-heating effects should include partial cooling of the cathode due to heat conduction through the thick LaS layers which was neglected in this study. The latter would allow to extend slightly the dc biasing operating range of the cathode beyond the estimate reported here. Furthermore, the thermal and electrical models of the cathode described here should be solved self-consistently. Once all these effects are taken into account, we believe the quantitative operation of the cold cathode exposed here will stay essentially correct predicting a dynamical shift of the

work function of the thin semimetallic film of the same order of magnitude than the one reported here.

Our analysis provides the basic design rules to fabricate a new cold cathode with emission windows with a rectangular or circular geometry. The growth of the structure would require the epitaxial growth of the structure shown in Fig.1. As discussed above, the epitaxial growth of *InP/CdS* heterostructures has been reported in the literature in the past [9]. The deposition of epitaxial *LaS* thin films has not been reported, to the best of our knowledge. We believe, however, that the figures of merits of the various cold cathodes analyzed in this work are a strong incentive towards the experimental investigation of these devices. If successful, such an experimental effort would lead to big pay-offs with the design of highly efficient cold cathodes for large panel displays, IR image convertors and sensors, and active power devices in mobile and airborne electronic equipment for military, commercial, and private use.

ACKNOWLEDGMENT

This work was sponsored by the Air Force Office of Scientific Research, Boiling AFB, D.C and by Wright laboratory.

References

- [1] S. Iannazzo, Solid State Electronics, Vol. 36(3), 301 (1993).
- [2] I. Brodie and C. A. Spindt, Advances in Electronics and Electron Physics, Vol. 83, p.2 (1992).
- [3] B. K. Williams and R. E. Simon, Physical Review Letters, Vol. 18(13), 485 (1967).
- [4] A. I. Akinwande, P. P. Ruden, B. L. Goldenberg, D. K. Arch, and John King, Proceedings of 1994 Tri-Service (NASA) Cathode Workshop, Cleveland, Ohio, March 1994, p. 105.
- [5] A preliminary account of this work has been published in: P. D. Mumford and M. Cahay, Journal of Applied Physics, Vol. 79(5), 2176 (1996). This work follows suggestions which are documented in two in-house reports by W. Friz, Final Technical Report, Task ELM-

9, June 1992, and Technical Note, Task ELM-6, January 1995 (Wright-Patterson Air Force Base).

- [6] The most desirable form of electron injection is accomplished by an ohmic contact because it puts minimal electrical stress on the material and, at the same time, provides the highest current densities.
- [7] Deposition of crystalline layers of *CdS* on *InP* by MBE are presented by W. G. Wilke, R. Seedorf, and K. Horn, *J. Vac. Sci. Technology. B*, Vol.7, p.807 (1989).
- [8] Chemical bath deposition (CBD) of crystalline layers of *CdS* on *InP* are discussed by D. Lincot, R. Ortega-Borges, and M. Froment, *Appl. Phys. Letters*, Vol. 64, p.569 (1994).
- [9] W. P. Shen and H. S. Kwok, *Appl. Phys. Letters*, Vol. 65(17), p.2162 (1994). In this article, the authors report the epitaxial growth of various II-VI compound semiconductors (*ZnS*, *ZnSe*, *CdS*, *CdSe*, and *CdTe*) on (111) and (100) *InP* and *GaAs* substrates by excimer laser ablation. All the films have good crystalline quality (fully in-plane) and mirror-like surface morphology. They found that, on (111)-oriented substrates, *CdS* and *CdSe* films were in the hexagonal phase with the *c* axis perpendicular to the surface, while *ZnS* and *ZnSe* films were in the cubic phase. The films grown on (100)-oriented substrates were all cubic. See also W. P. Shen and H. S. Kwok in *Compound Semiconductor Epitaxy*, edited by C. W. Tu, L. A. Kolodziejski, and V. R. McCrary, [*Mater. Res. Soc. Symp. Proc.* 340, 1994].
- [10] P. R. Bell, *Negative Electron Affinity Devices*, Oxford: Clarendon Press, 1973
- [11] Sze, *Physics of Semiconductor Devices*, 2nd Edition, Wiley, p. 291 (1981).
- [12] Assuming that the most part of the wide bandgap semiconductor is an intrinsic material (See Fig. 1), accumulation and depletion effects can be ignored. Referring to Fig. 1, the resulting uniform electric field across the wide bandgap semiconductor is then given by

$$E = \frac{1}{L_1} [V_{bias} + \Delta - (\frac{E_G}{2} - |\Delta\chi|)]. \quad (46)$$
- [13] Both T_1 and T_2 should actually be function of the number of reflections in the semimetallic slab since the electrons loose energy during each traversal. This is however just a refinement to the general theory exposed in this paper.

- [14] N. W. Aschcroft and N. D. Mermin, Solid State Physics, Saunders College, Philadelphia, 1976.
- [15] The room temperature work function for LaS was calculated by extrapolating measured work function values at high temperature as reported by S. Fomenko in *Handbook of Thermionic properties* (Plenum, New York, 1966). Within the range of temperature investigated by Fomenko, the LaS work function increases with temperature at a rate of $2meV/K$.
- [16] There are other reports for the electron affinity of CdS , including $\Delta\Xi_{CdS} = 4.5$ eV in [11]. In this case, Anderson's rule would allow electrons to spill over from the InP to the CdS region, an injection mechanism hard to control. The problem of the InP/CdS interface deserves more attention [9].
- [17] Shyh Wang, Fundamental of Semiconductor Theory and Device Physics, Prentice Hall Series in Electrical and Computer Engineering, Leon O. Chua, Series Editor (1989).
- [18] There are conflicting reports for the measured electron affinity of CdS . The affinity is known to vary with the surface orientation and the phase of the semiconductor material. Here, we use the electron affinity $\chi_{CdS} = 4.2$ eV reported in O. Madelung, "Semiconductors other than Group IV elements and III-V compounds ", Springer (1992).
- [19] E. S. Kohn, Journal of Applied Physics, Vol. 42(6), 2493 (1971).
- [20] M. Cahay, Final report, Summer Faculty Research Program, AFOSR, September 1996.
- [21] J. A. Edminister, Electromagnetics, Schaum's Outline Series, McGraw-Hill Book Company (1979).
- [22] M. L. Huberman and J. Maserjian, Physical Review B, Vol. 37(15), 9065 (1988).
- [23] R. C. Jaklevic and J. Lambe, Phys. Rev. B, Vol. 12, 4146 (1975).
- [24] Y. Silberberg and T. Sands, IEEE Journal of Quantum Electronics, Vol. 28 (7), 1663 (1992).

Table I: Material Parameters of the Cold Cathode

Material	Au (Ag)	n^{++} -InP	i-CdS	LaS
Lattice Thickness (\AA)	optional	optional	300.0	24.6
Lattice Constant (\AA)	4.04 (4.09)	5.86	5.83	5.85
Workfunction (eV)	4.3 (4.3)	4.4	4.2	1.14
Bandgap (eV)		1.42	2.5	
# of free electrons (10^{22}cm^{-3})	n^{++}	5.9 (5.86)	—	1.99
Electron Mass (m_0)	1.0	0.0765	0.14	1.0
Electron Mobility ($\text{cm}^2 \text{V}^{-1} \text{s}^{-1}$)		5370.	400.0	
Thermal conductivity @ 300K (W/cmK)	3.1 (4.18)	0.74	0.05..1	0.17
Electrical resistivity (273K) ($\mu\Omega \text{cm}$)	1.51 (2.04)			92.0
Melting temperature (K)		1335.		2500

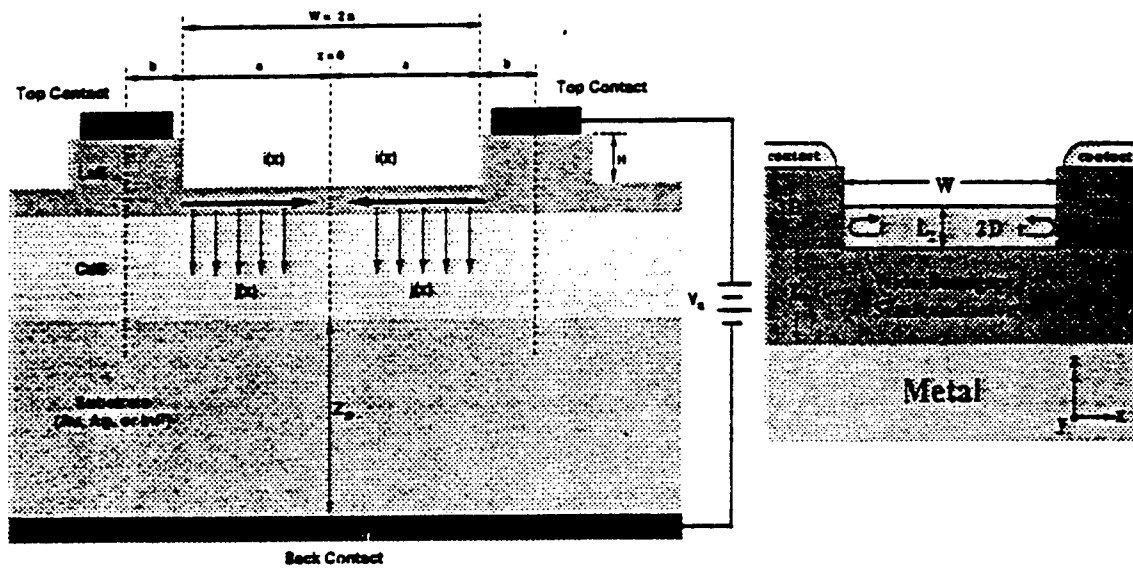
Table II: Physical Parameters of the Cold Cathode

Thickness of InP substrate	50 - 200 μm
Thickness of CdS thin film	300 - 500 \AA
Emission window length	1 cm
Thickness of LaS thin film	24.6 \AA
Thickness of LaS thick regions	100 - 500 \AA
Electron mean free path in LaS thin film	50 - 300 \AA

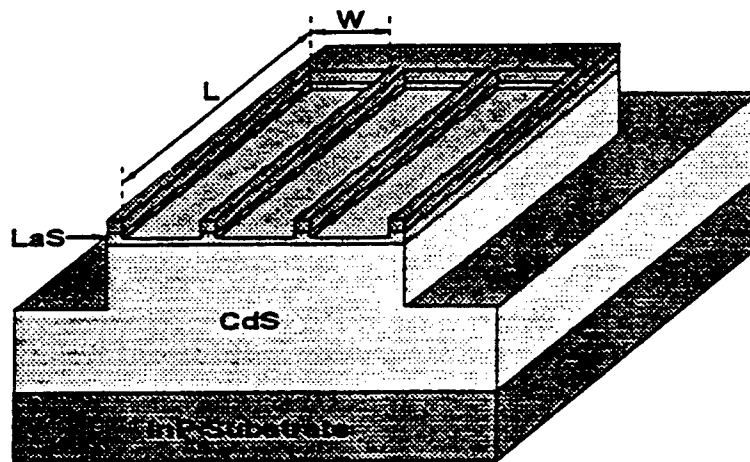
Table III: Parameters for numerical fit (Eq.(10)) to the Fowler-Nordheim current expression (Eq.(9)). The fit is made over a current density range from 1 to 1000 A/cm².

Width of Rectangular Window (μm)	J_0 (A/cm ²)	α (V ⁻¹)
10,000	7.2186×10^{-7}	2.3651
1,000	3.6483×10^{-6}	2.1302
500	3.8505×10^{-6}	2.1224
200	4.7337×10^{-6}	2.0933
100	4.7902×10^{-6}	2.0916
50	4.8207×10^{-6}	2.0907
20	4.8424×10^{-6}	2.0901
10	4.8424×10^{-6}	2.0901

Figures

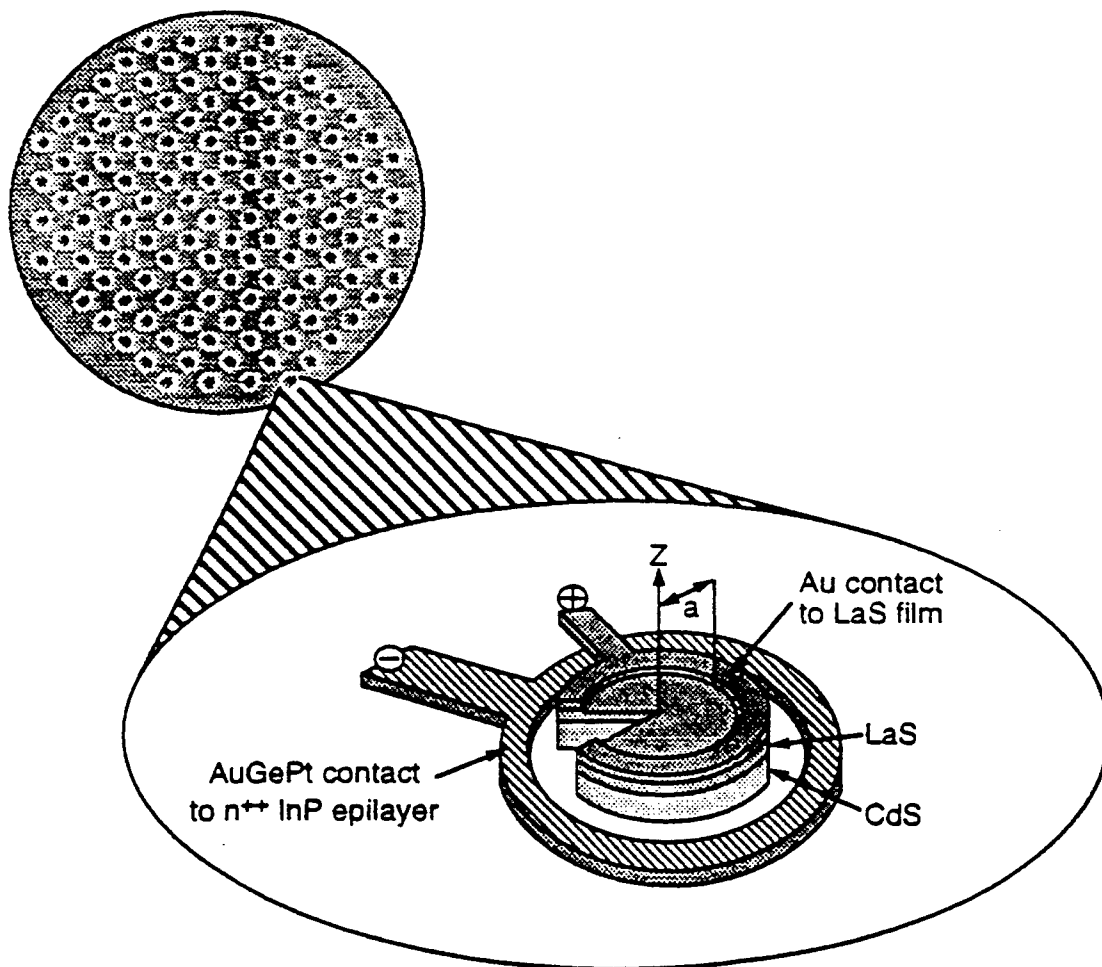


(a)



(b)

Figure 1: (a) Left: cross-section of the newly proposed cold cathode between two emitter fingers. Trapping of electrons by the *LaS* semimetallic thin film leads to a lateral current flow and current crowding in the structure. Right: illustration of the partial reflection of the two-dimensional electron gas in the *LaS* thin film upon entering the three-dimensional contact regions where the external bias is applied to *Au* contacts made to the thick *LaS* regions. (b) Illustration of the multiple finger metallic structure used to bias appropriately a cold cathode with rectangular emission window.



(c)

Figure 1 (cont'd): (c) Array of circular cold cathode emitters arranged in a honeycomb configuration and close-up view of one of the emitters showing the various layers in the epitaxially grown structure and the contacts made to the *LaS* metallic grid and the $n^{++} - \text{InP}$ heavily doped injection layer. [After W. Friz, Task ELM-2, "Dissipative Processes in Veiled Work Function Emitters", Contract F 33615-C-95-1755, WPAFB, Dayton, May 1996].

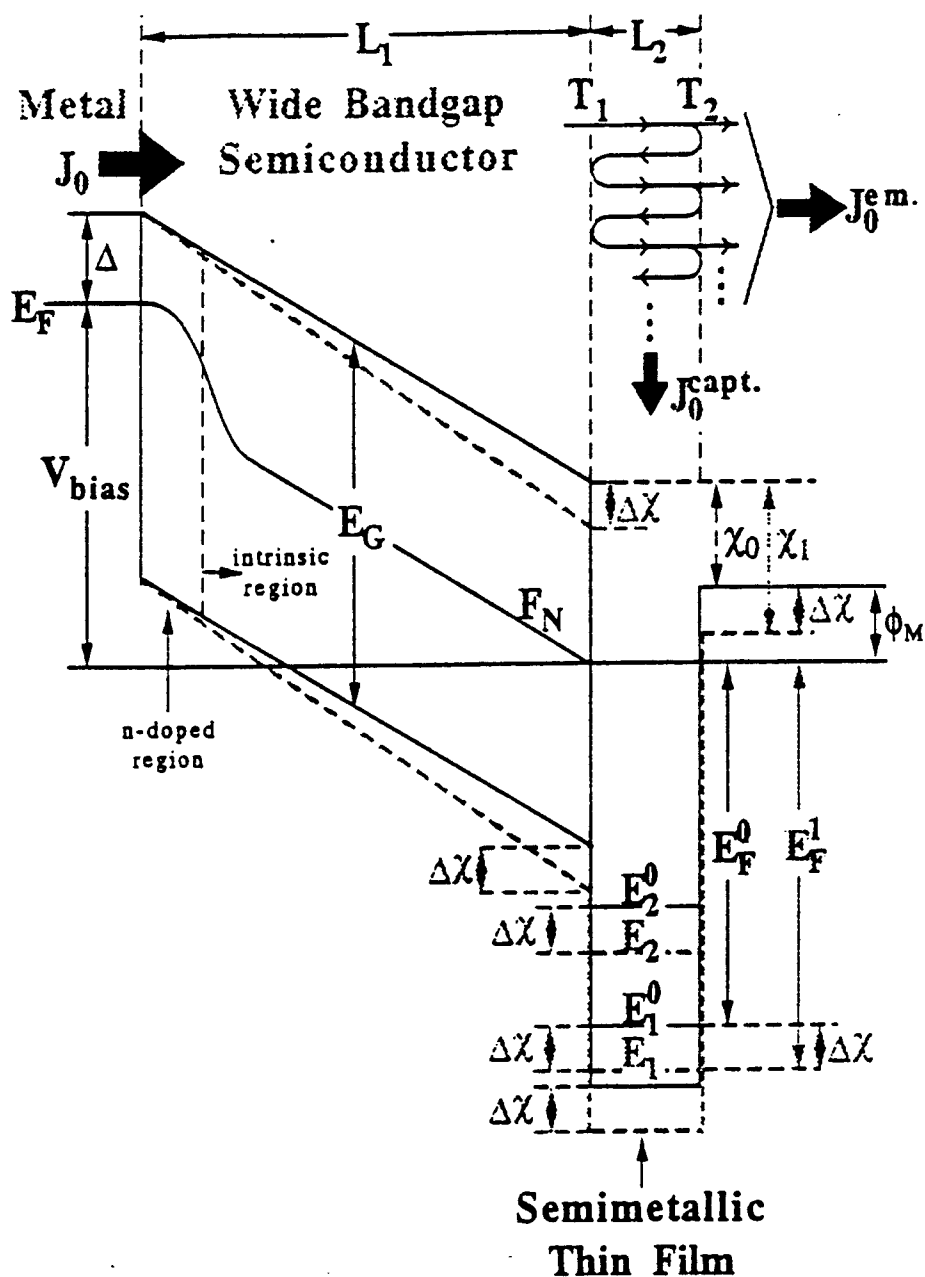


Figure 2: Schematic representation of the conduction band profile throughout the cold cathode emitter described in the text. Under forward bias, a fraction of the emitted current is captured in the semimetallic slab. The subsequent excess sheet carrier concentration in the quantum well formed by the semimetallic slab leads to a shift of the fermi level in the thin film which is similar to a lowering of the work function of the thin film. For a given forward bias, this leads to an increase in the electric field in the wide bandgap semiconductor (dashed line versus full line) and in an increase in the injection and emitted currents. Also shown in the quasi Fermi level spatial dependence across the wide bandgap semiconductor.

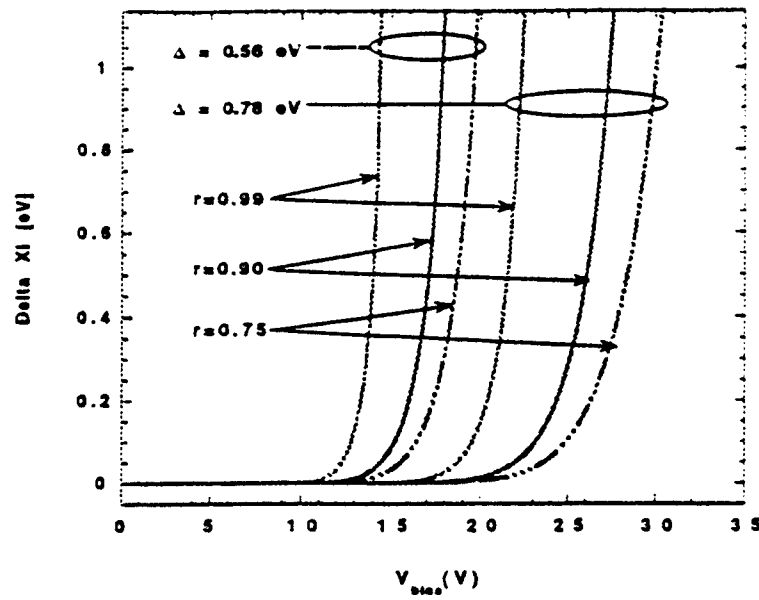


Figure 3: Dynamical shift of the work function as a function of the external applied bias for a cold cathode emitter with the parameters listed in Table I. For each group of curves, $r = 0.99$, 0.9 , and 0.75 , from left to right. The following physical parameters were used: $\lambda_{LaS} = 300 \text{ \AA}$, $T_1 = T_2 = 0.5$, and $v_F = 1.36 \times 10^8 \text{ cm/s}$.

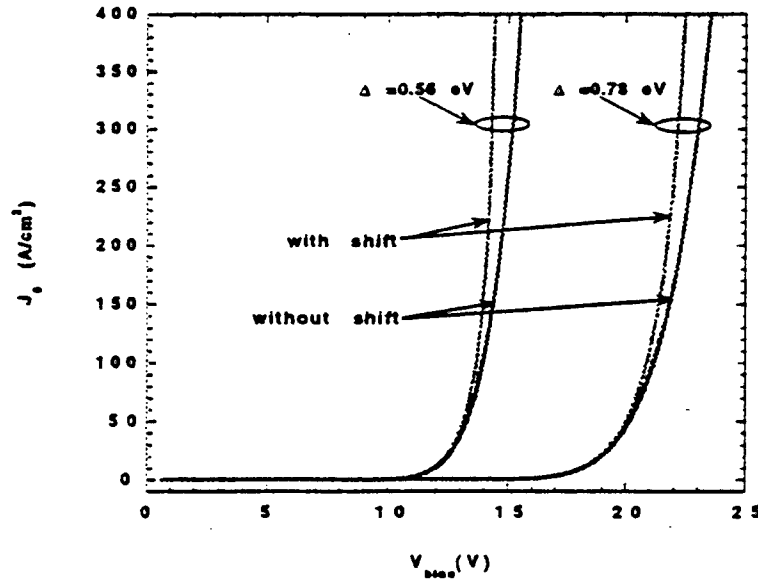
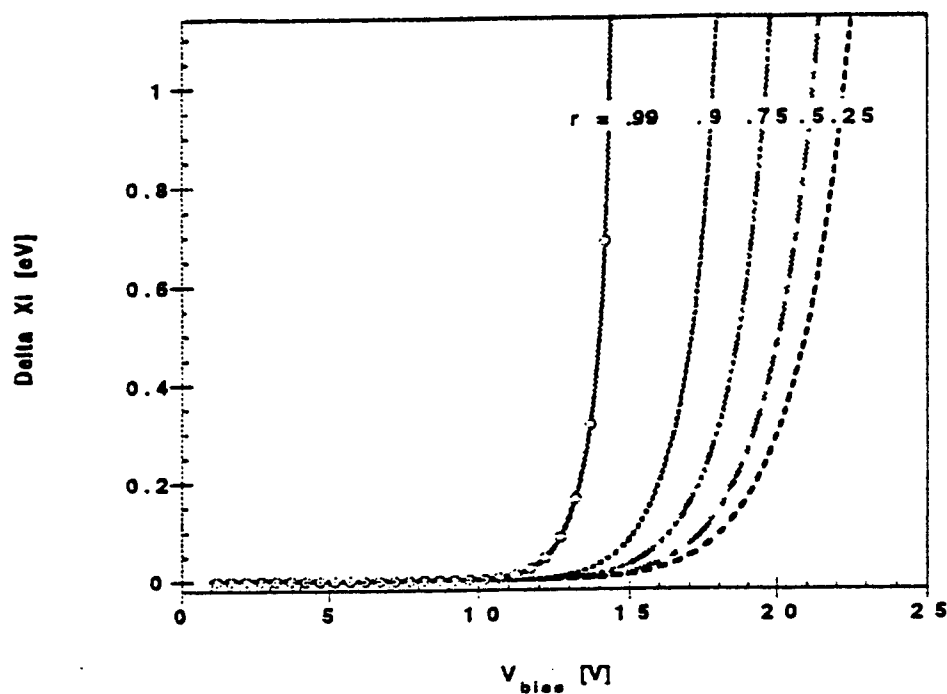
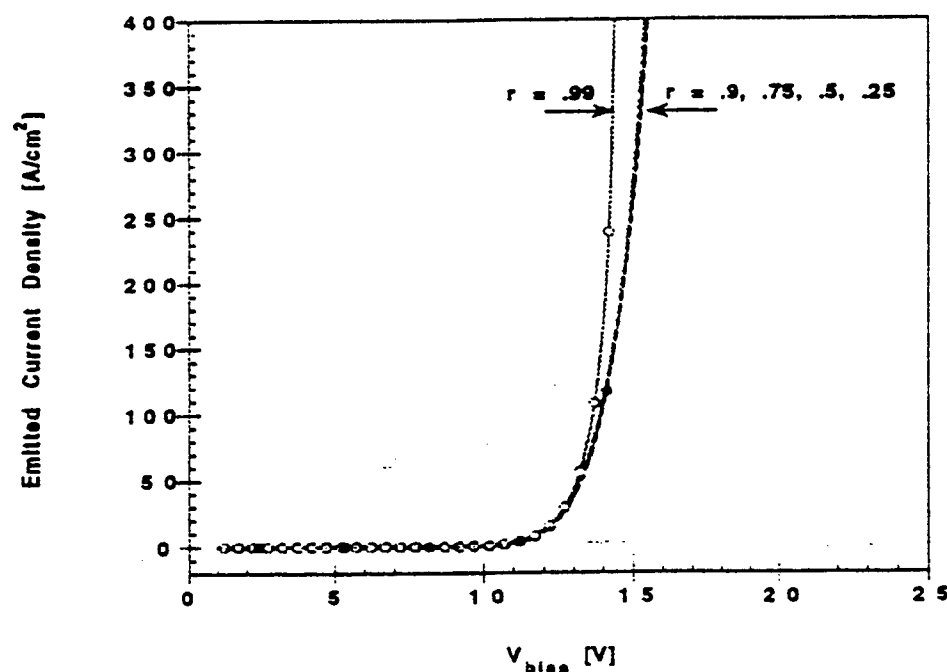


Figure 4: Comparison between the emitted current as a function of the applied bias while including (dashed line) and neglecting (full line) the dynamical shift of the work function of the semimetallic slab described in the text. The parameters of the device are listed in Table I. The coefficient C_1 and C_2 in Eq.(1) were chosen equal to $1.5 \times 10^6 \text{ A/V}$ and $6.9 \times 10^7 \text{ (V}^{1/2} \text{ cm)}^{-1}$, respectively. Other choices for the parameters C_1 and C_2 with similar magnitudes lead to similar dynamical shift of the work function. The reflection coefficient r shown in Fig. 1(a) was set equal to 0.99 . The following parameters were used: $\lambda_{LaS} = 300 \text{ \AA}$, $T_1 = T_2 = 0.5$, and $v_F = 1.36 \times 10^8 \text{ cm/s}$.

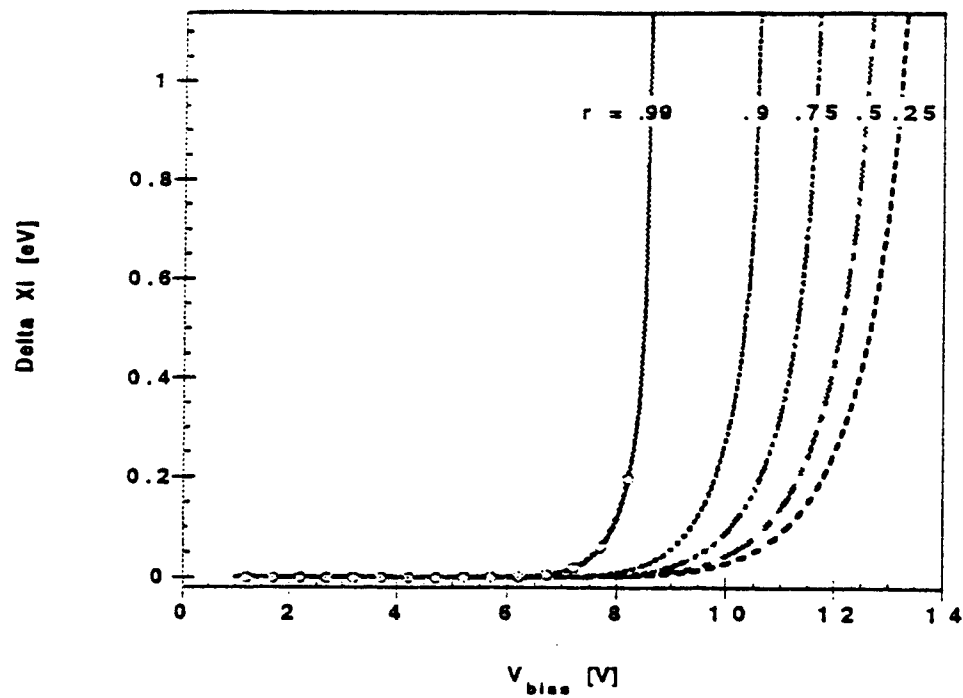


(a)

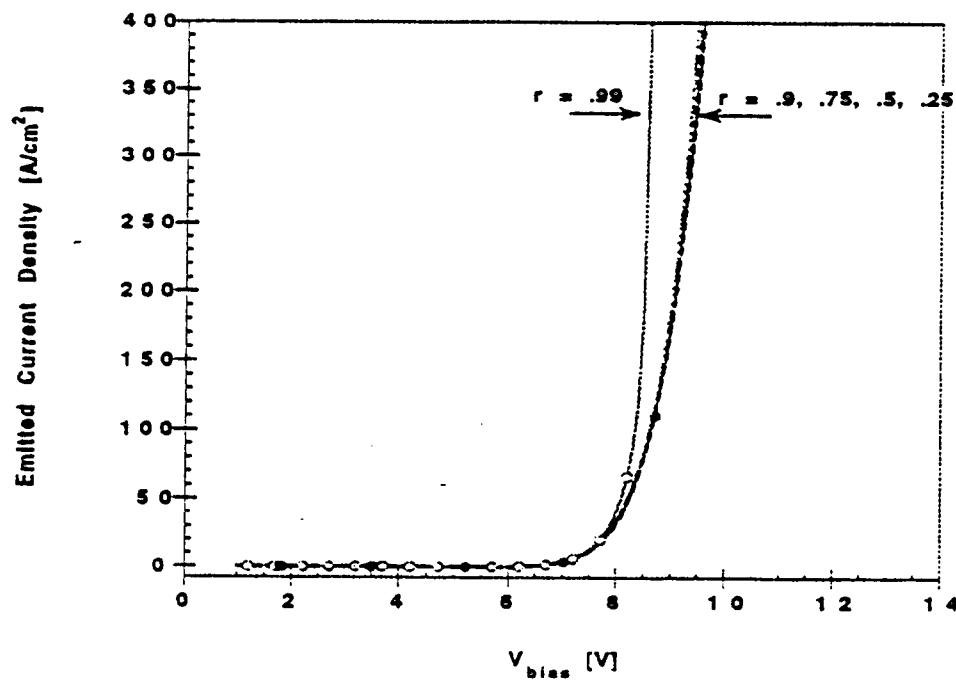


(b)

Figure 5: (a) Bias dependence of the dynamic work function shift in a typical cold cathode while varying the reflection coefficient of electrons between the 2D thin film and LaS contact regions. (b) Corresponding bias dependence of emitted current density. The following parameters were used ($L_1 = 500 \text{ \AA}$, $L_2 = 24.6 \text{ \AA}$, $W = 1 \text{ cm}$, $T_1 = T_2 = 0.5$, $\lambda_{LaS} = 300 \text{ \AA}$, $\Delta(A_g) = 0.56 \text{ eV}$).

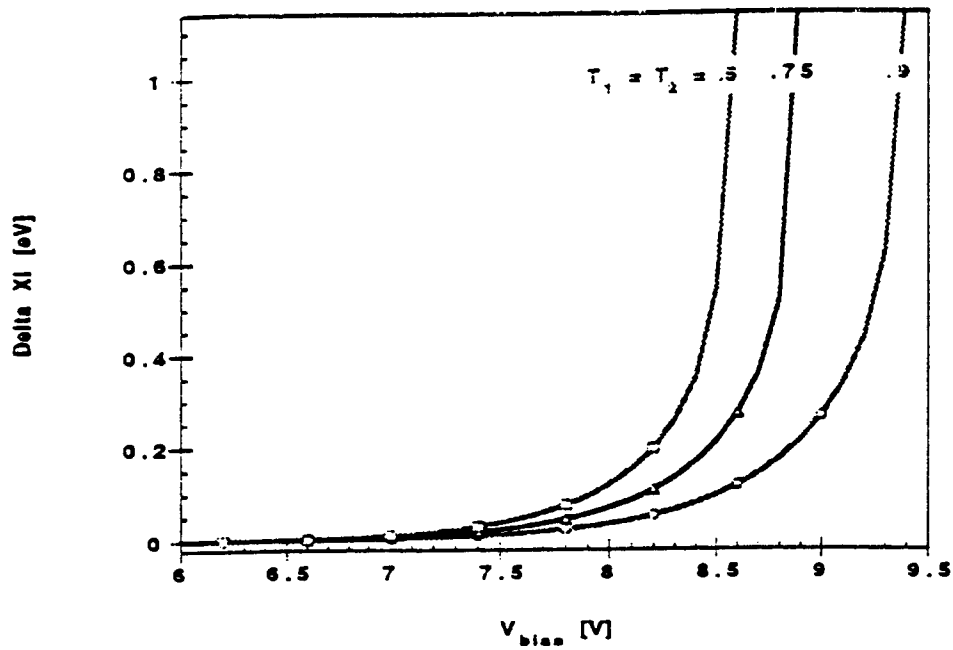


(a)

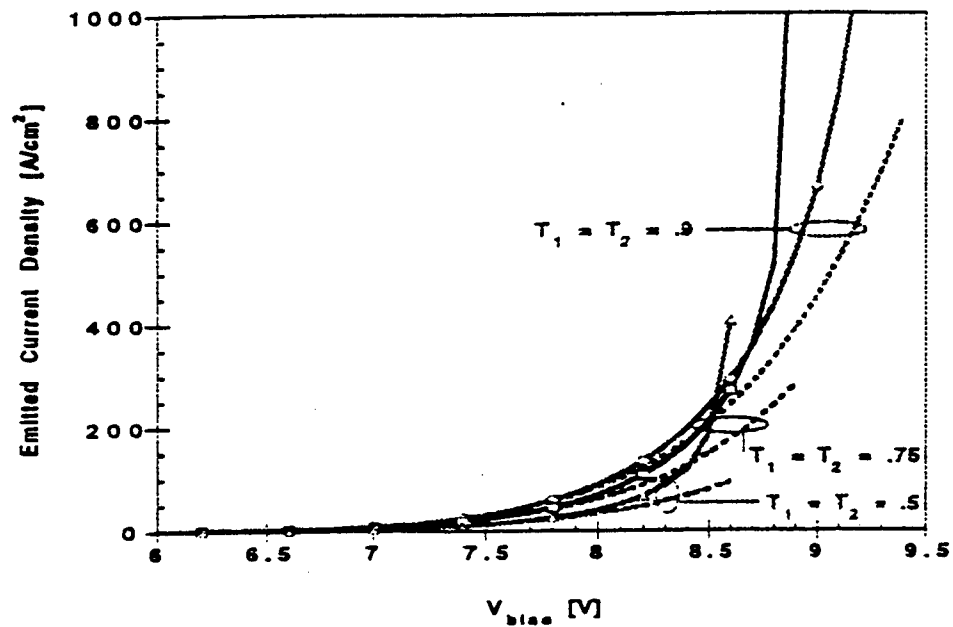


(b)

Figure 6: Same as Figure 5 for a cathode with a 300 Å thick *CdS* region, all other parameters being kept the same. Comparison with Figure 5 shows that the exponential rise of the dynamic work function shift and the emitted current density occurs at smaller values of the applied bias.

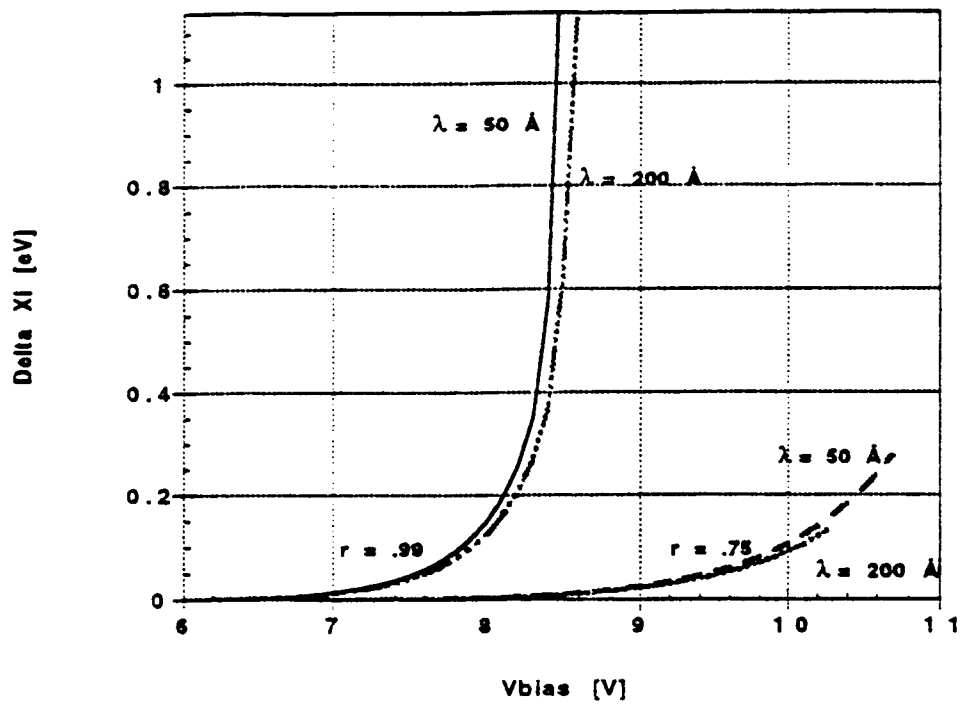


(a)

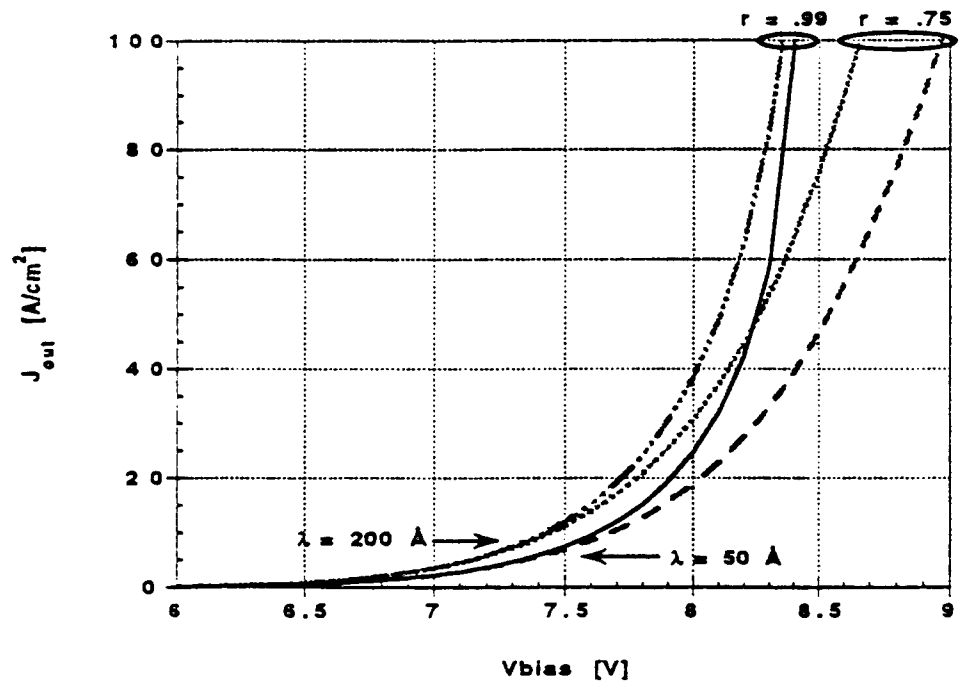


(b)

Figure 7: (a) Bias dependence of the dynamic work function shift in a typical cold cathode while varying the transmission coefficients T_1 and T_2 at the CdS/LaS and $LaS/vacuum$ interfaces. (b) Corresponding bias dependence of emitted current density. The full and dashed lines are the current density plots calculated with and without the effect of the dynamic work function shift. The following parameters were used ($L_1 = 300 \text{ \AA}$, $L_2 = 24.6 \text{ \AA}$, $W = 1 \text{ cm}$, $\lambda_{LaS} = 300 \text{ \AA}$, and $\Delta(A_g) = 0.56 \text{ eV}$). The curves are labelled with the values of the transmission probabilities T_1 and T_2 which were assumed to be identical.

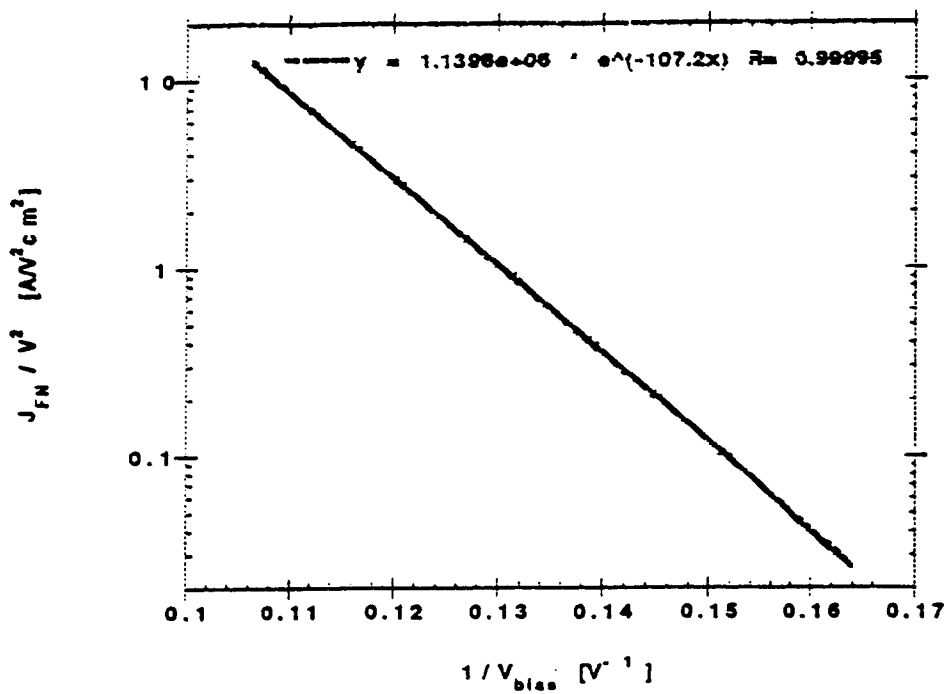


(a)

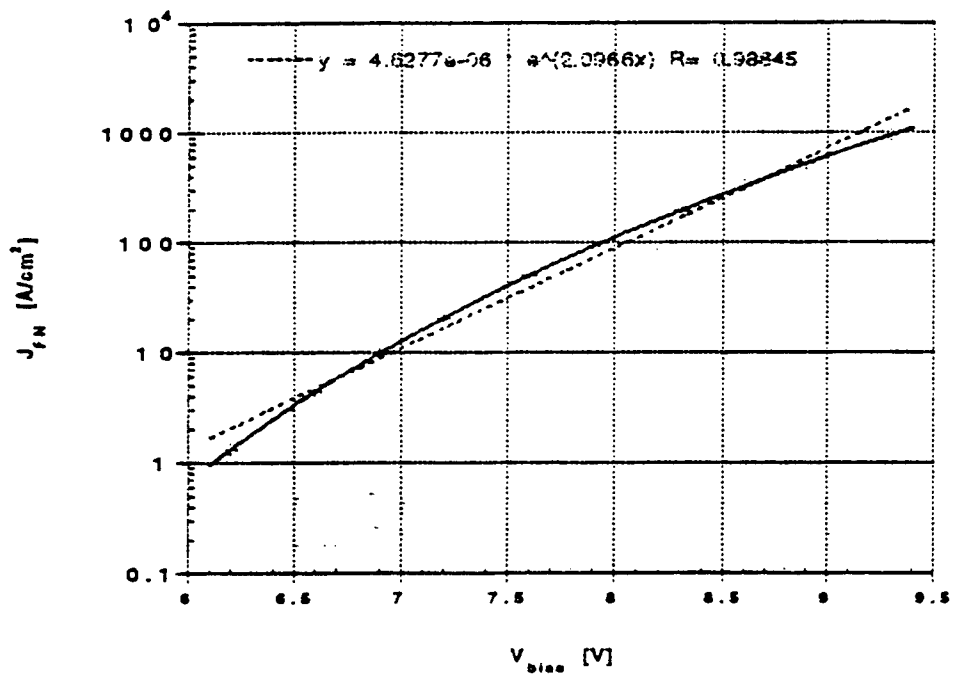


(b)

Figure 8: (a) Bias dependence of the dynamic work function shift in a typical while varying the electron mean free path λ_{LaS} . (b) Corresponding bias dependence of emitted current density. The following parameters were used ($L_1 = 300 \text{ \AA}$, $L_2 = 24.6 \text{ \AA}$, $W = 1 \text{ cm}$, $\lambda_{LaS} = 300 \text{ \AA}$, $\Delta(A_g) = 0.56 \text{ eV}$, $T_1 = T_2 = 0.5$, and $r = 0.99$).



(a)



(b)

Figure 9: (a) Fowler-Nordheim fit to the bias dependence of the emission current of a cold cathode with the physical parameters listed in the caption of Figure 5. (b) Fit of the same J_{FN} versus V_{bias} calculation by an expression of the type $J_{FN} = J_0 e^{\alpha V_{bias}}$ for a range of V_{bias} for which the emitted current density varies between 1 and $1000 A/cm^2$.

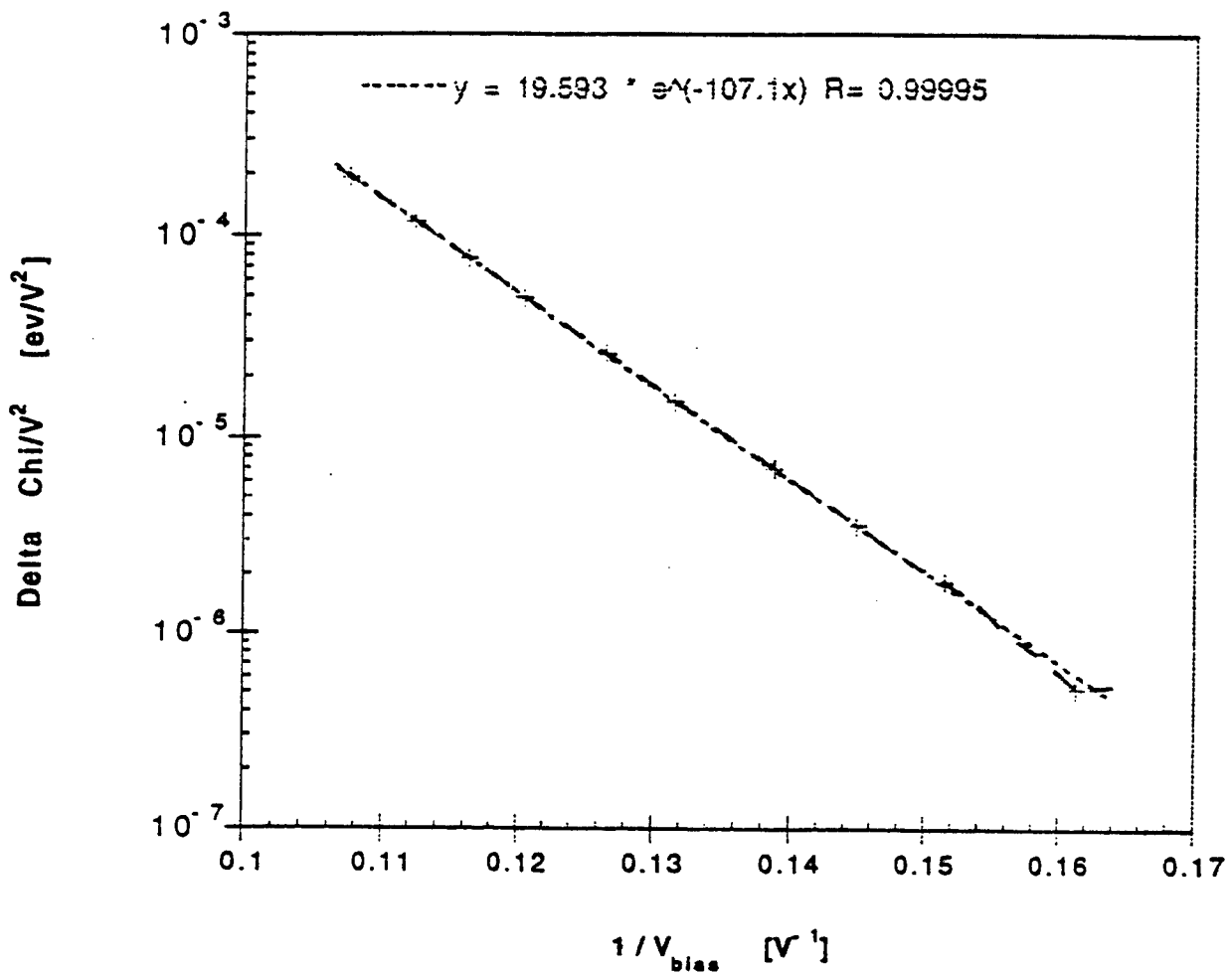


Figure 10: Fowler-Nordheim fit to the bias dependence of the dynamic work function shift for a cold cathode with the physical parameters listed in the caption of Figure 5. The following parameters were used ($L_1 = 300 \text{ \AA}$, $L_2 = 24.6 \text{ \AA}$, $W = 1 \text{ cm}$, $\lambda_{\text{LAS}} = 300 \text{ \AA}$, and $\Delta(A_g) = 0.56 \text{ eV}$).

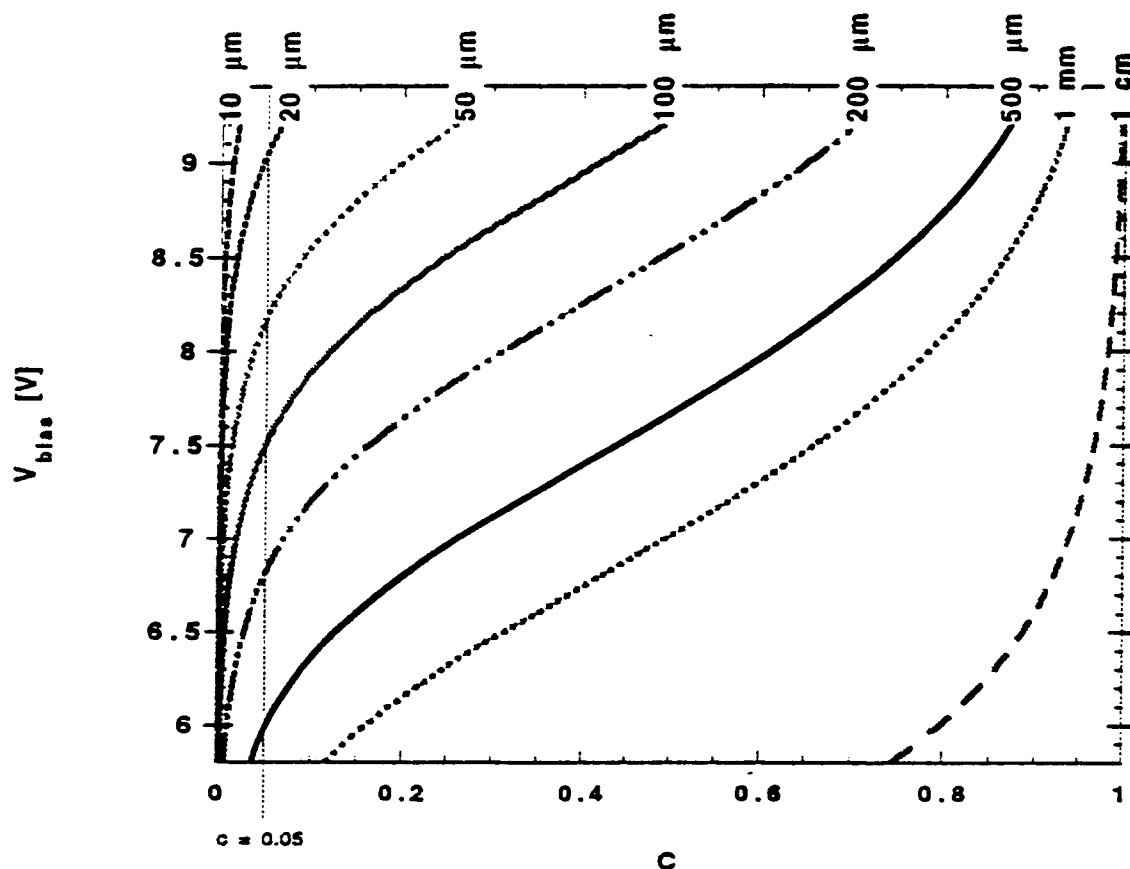


Figure 11: Bias dependence of the parameter c as a function of applied bias for different radii of a cold cathode emitter with a circular geometry. Cold cathodes with narrower emitter window must be operated over a smaller dc bias range to avoid current crowding. The following parameters were used ($L_1 = 300 \text{ \AA}$, $L_2 = 24.6 \text{ \AA}$, $W = 1 \text{ cm}$, $\lambda_{LaS} = 300 \text{ \AA}$, $\Delta(A_g) = 0.56 eV$, and $r = 0.99$). Current crowding effects are negligible as long as the parameter c is kept under 0.05.

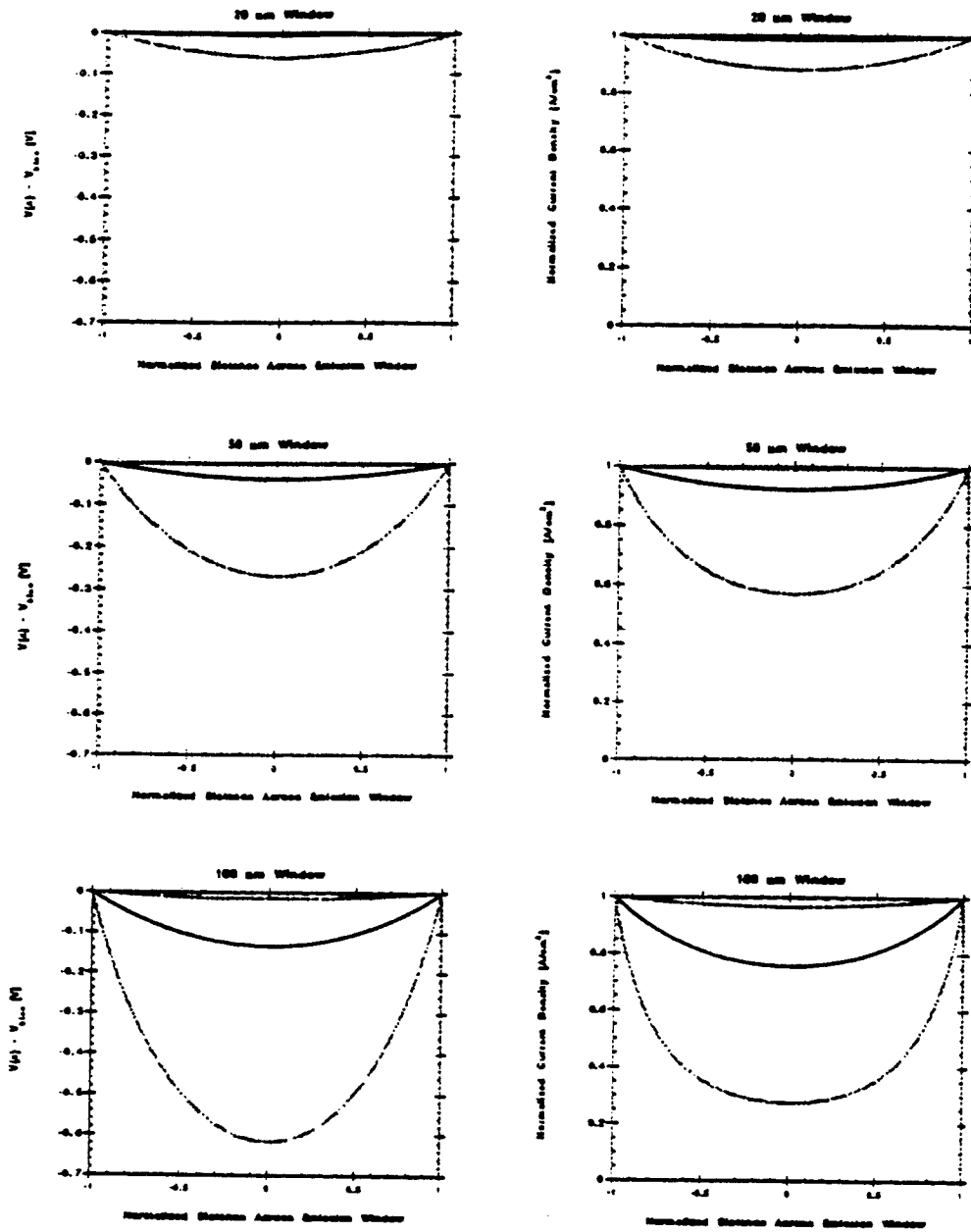


Figure12 : Radial variation of the potential drop across the LaS windows of various radii. The cathode is assumed to operate at room temperature and the electrical resistivity of LaS is set equal to $92 \mu\Omega cm$. The potential is measured relative to the potential at the rim of the LaS circular window where the electrostatic potential is assumed to be equal to the applied bias. For each figure, the values of the normalizing current density are from top to bottom 1, 10, 100, 1000 A/cm². The LaS window radius is indicated on top of each figure. The following parameters were used ($L_1 = 300 \text{ \AA}$, $L_2 = 24.6 \text{ \AA}$, $W = 1cm$, $\lambda_{LaS} = 300 \text{ \AA}$, and $\Delta(A_g) = 0.56eV$). Right frames : corresponding radial dependence of the emitted current densities. Current densities are normalized to the values of the emitted current at the window rim.

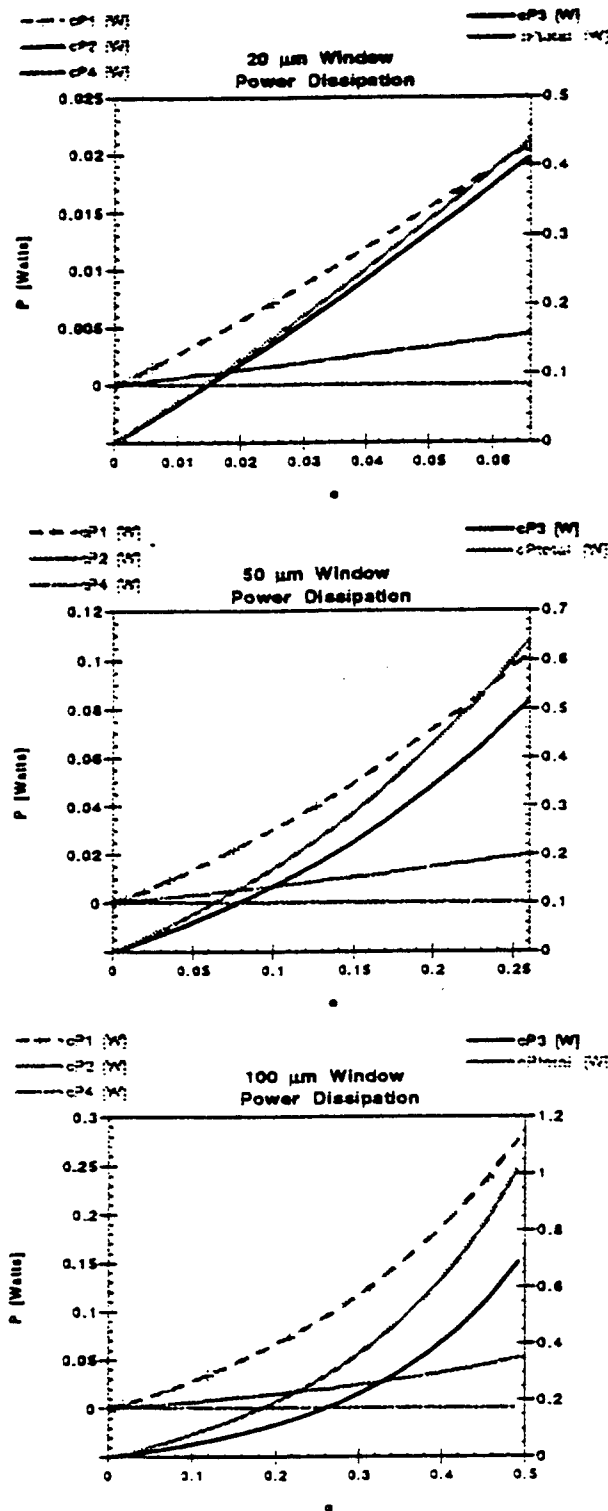
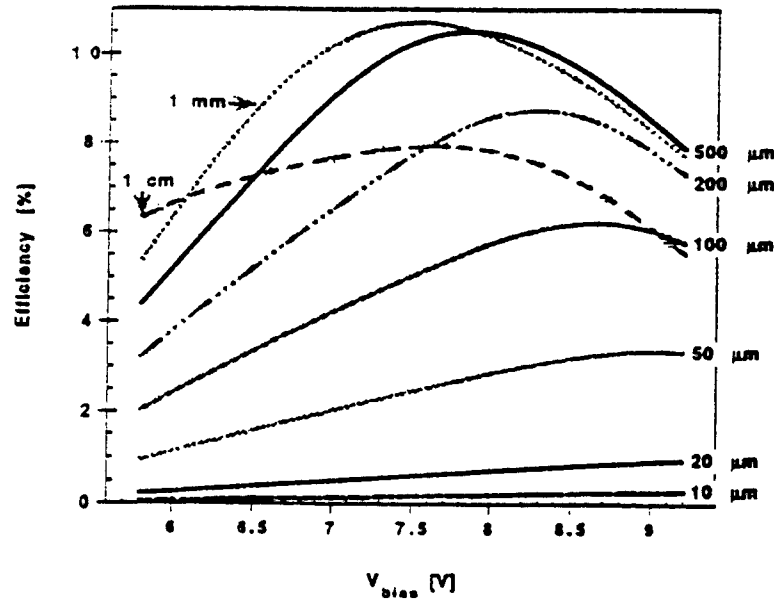
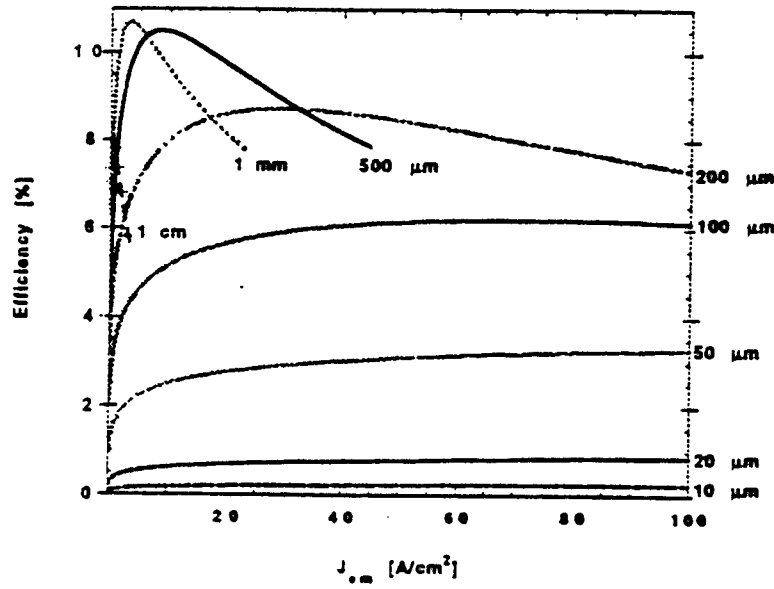


Figure 13 : Variation with the parameter c of the different contributions to the total power dissipated in circular cold cathodes with different radii (see Eqns.(32-39)). The physical parameters of the cathodes are listed in Table I and in the caption of Fig.5. The CdS layer is assumed to be grown on a heavily doped InP substrate $100\mu\text{m}$ thick. The thermal conductivity of the InP substrate was taken equal to 0.74 W/Kcm and the parameter b in Eq.(13) was set equal to 1.45 . The thickness of the LaS contact regions was set equal to 100 \AA .



(a)



(b)

Figure 14: Power efficiency η_P of cathodes of different radii is plotted as a function of (a) V_{bias} and (b) the emitted current density J_{em} . For all window sizes, the efficiency decreases with V_{bias} and J_{em} as a result of current crowding. The efficiency is more or less constant over a wider range of V_{bias} for window with smaller radius because current crowding is less important in that case, as illustrated in Fig.12 . The overall lower efficiency for smaller size window comes from the fact that the width b of the thick LaS regions was set equal to $100\mu m$ for all cathodes. The efficiency of cathodes can be increased by making the ratio b/a in Fig.1 closer to unity. The following parameters were used ($L_1 = 300 \text{ \AA}$, $L_2 = 24.6 \text{ \AA}$, $W = 1\text{cm}$, $\lambda_{LaS} = 300 \text{ \AA}$, $\Delta(A_g) = 0.56\text{eV}$, $T_1 = T_2 = 0.5$, and $r = 0.99$).

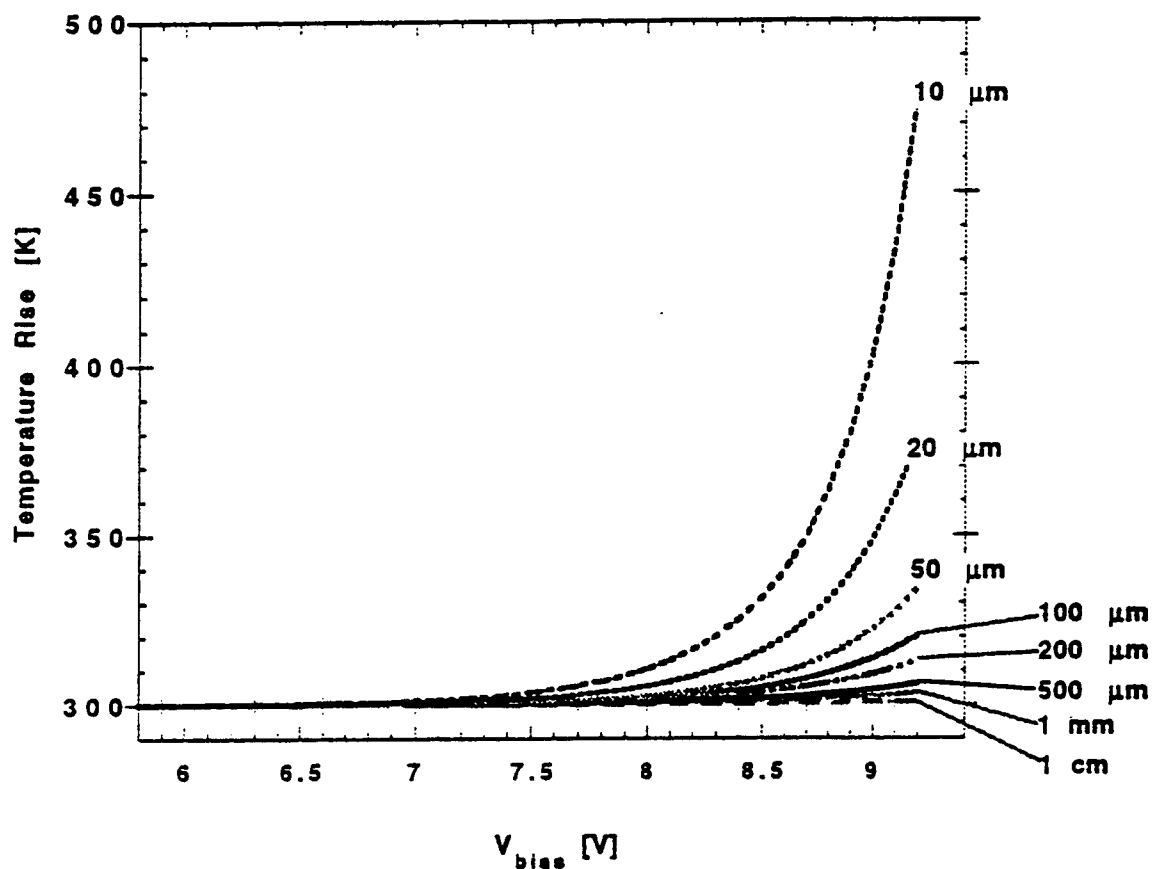


Figure 15: Temperature rise in the *LaS* thin film as a function of the applied bias for the various cold cathode structures studied in Figure 13. The active area of the cathode (CdS/*LaS* layers) is assumed to be grown on a 100 μm thick InP substrate with the back of the substrate acting as a perfect heat sink (300K).

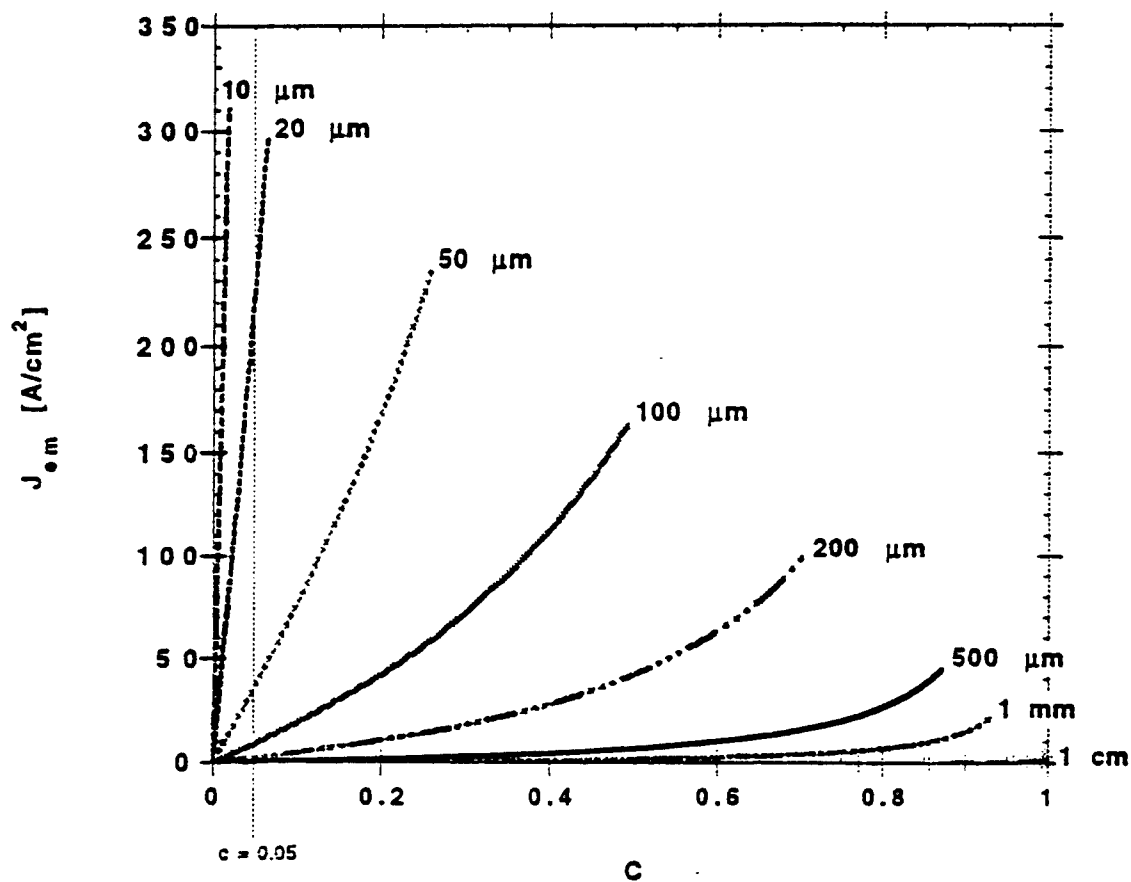


Figure 16 : Emitted current densities for emitter window of different radii as a function of the parameter c . The vertical line at $c=0.05$ is the line to the right of which the effects of current crowding can not be neglected. Appreciable emitter current densities can be obtained for cathodes with emitter window width under $50\mu m$. The following parameters were used ($L_1 = 300 \text{ \AA}$, $L_2 = 24.6 \text{ \AA}$, $W = 1cm$, $\lambda_{LaS} = 300 \text{ \AA}$, $\Delta(A_g) = 0.56eV$, $T_1 = T_2 = 0.5$, and $r = 0.99$).

**A NOVEL COMPATIBILITY/EQUILIBRIUM BASED ITERATIVE
POST-PROCESSING APPROACH FOR AXISYMMETRIC
BRITTLE MATRIX COMPOSITES**

Reaz A. Chaudhuri, Associate Professor
Department of Materials Science & Engineering,
304 EMRO Building

University of Utah
Salt Lake City, UT 84112

Final Report for:
Summer Faculty Research Program
Wright Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC
and
Wright Laboratory
Wright -Patterson Air Force Base, Ohio

March, 1998

A NOVEL COMPATIBILITY/ EQUILIBRIUM BASED ITERATIVE POST-PROCESSING APPROACH FOR AXISYMMETRIC BRITTLE MATRIX COMPOSITES

Reaz A. Chaudhuri, Associate Professor

Department of Materials Science & Engineering, University of Utah

Abstract

A semi-analytical iterative approach for enhancing the existing two-dimensional quasi-continuous axisymmetric stress field for a brittle matrix micro-composite (i. e., a single fiber surrounded by a concentric matrix cylinder), is presented. The existing solution employs Reissner's variational theorem in conjunction with an equilibrium stress field in which the radial (r -) dependence is assumed *a priori*.

In the present approach, the stress distribution in the radial direction obtained from the afore-cited variational model is improved *a posteriori* through an iterative approach that involves successive substitution of the previously computed strains (or stresses) into the equations of compatibility and equilibrium. The equations of compatibility are selected such that they form Euler equations corresponding to appropriate variational principle, such as the principle of minimum complementary potential energy, etc. The boundary/interface conditions at $r = \text{constant}$ and $z = \text{constant}$ surfaces/interfaces are satisfied in the pointwise sense. The expressions for the improved axisymmetric displacement and stress fields are derived using the symbolic language, MAPLE. An illustrative thermal stress problem is currently being solved, and will be used to compare with the existing variational solution.

1. Introduction:

Studies of the behavior of unidirectional and laminated composites made from stiff elastic matrix materials which may develop imperfect interfaces with the fibers have enjoyed a revival after the early classical work of Aveston, Cooper, and Kelly (1971). Pagano (1991) refers to these materials as brittle matrix composites (BMC). The aforementioned ACK modeling, as well as the more recent development proposed by Budiansky, Hutchinson, and Evans (1986), are based upon primitive approximations of the stress field developed within a concentric cylinder, i.e., a circular cylindrical body of one material surrounded by a concentric annulus or ring of a second material. There exists a considerable body of literature associated with the elasticity problem of a concentric cylinder, where modern interest is focused on its use as a representative volume element (RVE) of a unidirectional composite (Hashin and Rosen, 1964; Pagano and Tandon, 1988). Additionally, a significant segment of the composite literature is based upon the one-dimensional shear lag analysis, which was apparently originated by Cox (1952). The results obtained using this kind of analysis are too inaccurate to merit further attention in this report.

Sternberg (1970) solved several axisymmetric load diffusion problems within a concentric cylindrical domain, in which the elasticity formulations are simplified by the assumption that the core material (fiber) can be modeled as a one-dimensional bar while the annular region (matrix) in all cases extends to infinity. An exact solution was derived for the case in which the bar was circular in cross-section, and was fully immersed within the unbounded matrix. A similar approach, with the difference of the fiber being assumed rigid, was employed by Luk and Keer (1979). This assumption is more appropriate for resin matrix composites, where the fiber to matrix modulus ratio is very high. The axisymmetric elasticity problem of a broken fiber embedded in an infinite matrix was treated by Pickett and Johnson (1967). In that work, the fiber is represented as a three-dimensional elastic medium; however, the report contained no numerical results for the stress field. Smith and Spencer (1970) also formulated an axisymmetric elasticity solution by a semi-inverse method for a class of boundary value problems in which the radius and length of the concentric cylinder are both finite. The solution is expressed in the form of a series of non-orthogonal functions which satisfy the field equations exactly. Homogeneous boundary conditions on the radial surface of the body are satisfied exactly while realistic end conditions can be approximated. The singularities predicted by Zak (1964) are smoothed out in this approach. A very formidable study, which includes correlation with experimental observations, is that by Atkinson et al (1982). In this work, the pullout of a single fiber from a matrix cylinder is treated. The fiber extends only partially along its length into the matrix. Results are provided for a perfectly bonded fiber-matrix interface as well as for states in which lateral (curved) surface debonding or fiber end plane debonding take place. The problem is solved by "patching" the asymptotic

singular stress field to that given by finite elements, although this method is not always reliable. An interesting conclusion is reached that, at least in a qualitative sense, the interface failure response can be anticipated from the stress field within the uncracked rod. In a model similar to Pagano's (1991), McCartney (1990) treated a class of concentric cylinder problems in which matrix cracking or debonding with or without friction are present. In that work, the functional r -dependence of the stress components is assumed which leads to a system of ordinary differential equations in z . All appropriate field equations are satisfied with the exception of two of the constitutive relations while some of the boundary/interface conditions could only be satisfied in an average sense. The ease and potential effectiveness of this model for composite analysis will demand its careful consideration in comparison with exact solutions and solutions given by competitive approaches. For example, Kurtz and Pagano (1991) formulated an infinite series solution of the axisymmetric elasticity problem in which a fiber is being pulled from the matrix. The length of the body as well as the outer radius are finite. Although the singularity is not explicit in the solution, Cesaro summation has been employed to improve the convergence of the stress field within the singular region.

The solution for the stated problem has been obtained by employing a modified version of the variational model (Pagano, 1991) of an axisymmetric concentric cylinder, which was successfully implemented earlier in the case of a flat laminate by Pagano (1978). The model is generated by subdividing the body into regions consisting of a core and a number of shells of constant thickness and length and satisfying the Reissner variational equation (1950) with an assumed stress field in each region. The number of regions, in particular in the r direction, can be increased in order to improve solution accuracy. The regions are selected such that the thermoelastic properties are constant and the boundary conditions do not change character on any of the bounding surfaces within each region. Pagano (1991) has sought to predict the influence of various kinds of damage (see Figure 1) and their interactions by accurately modeling the micro-mechanical stress field in their presence by using the afore-cited variational approach.

Strengths and Weaknesses of the Existing Approach (Pagano, 1991)

Strengths:

1. This assumed stress based approach insures satisfaction of the axisymmetric equilibrium equations;
2. The existing method utilizes a reasonably accurate non-singular axial (z -direction) variation of the computed stress field that also satisfies the end ($z = \text{const.}$) boundary conditions (i. e., no artificial discontinuity due to sectioning).

Weaknesses:

1. Layering in the radial direction, that introduces artificial discontinuities in some stress components at a layer interface within the same material;

2. As the number of layer increases, the computed eigenvalues become numerically very large, thus limiting the number of layers required for accurate stress field in the vicinity of a stress singularity point.

Objectives of the Current Research:

The present method seeks to alleviate the afore-mentioned weaknesses of the existing approach, while preserving its inherent strengths. The specific goals of the present investigation includes

- Improvement of the radial (r -) variation of the stresses so that subdivision into very thin layers and the associated blow-up of the computed eigenvalues can be avoided;
- Artificial discontinuities of the stress components, σ_θ and σ_z , at a layer interface within the same material can be avoided;
- Determination of the "stress intensity factor" by way of matching the local asymptotic stress field with the afore-cited improved solution at an arbitrarily close distance from the point of stress singularity (i. e., fiber-matrix interface located at a free edge).

As a first step, Chaudhuri (1996) presented an approximate "plane strain" version of the present solution, wherein the equations were simplified by dropping $1/r$ terms in the equilibrium equations. Additionally, the constitutive relations used were those due to the plane strain. The primary reason for resorting to this approximation was the fact that Chaudhuri's (1996) previously "improved" axisymmetric solution generated 10 "constants" of integration for a 2-layer fiber-matrix cylinder model against 8 interface/boundary conditions in the radial direction. Plane strain approximation removed 2 constants of integration, thus rendering the boundary-value problem under investigation solvable in closed form. However, although the preliminary results thus obtained demonstrated that the plane strain based approach can be implemented in the existing FORTRAN code due to Pagano and co-workers (1991), the accuracy of the results were far from encouraging. This discrepancy was due to the fact that although the plane strain condition prevails in the vicinity of the point of stress singularity (i. e., fiber-matrix interface located at a free edge) as was shown by Zak (1964), the same is not true in the far field where the boundary conditions are applied. The present (summer, 1997) research has corrected this situation by rederiving the correct axisymmetric solution with 4 "constants" (i.e., functions of z) of integration per layer. This is currently being implemented in the framework of symbolic language MAPLE software, for the purpose of solving a system of simultaneous ODE's in terms of the integration "constants" (functions of z).

Secondly, although the boundary/interface conditions at $r = \text{constant}$ surfaces were satisfied in the pointwise sense, the past (summer, 1996) approach left the end boundary conditions in the axial direction (i.e., at surfaces $z = \text{constant}$) undefined thus rendering the boundary-value formulation ill-posed. This ill-posedness was removed by introducing boundary error terms that helped satisfy the end boundary conditions at surfaces $z = \text{constant}$ in

a somewhat ad hoc fashion. The present (summer, 1997) research has identified a novel approach wherein the z-boundary conditions, that can be satisfied in a pointwise sense, are generated using the equations of the theory of elasticity within a particular cycle or iteration. However, this "discovery" came too late to be implemented in the framework of MAPLE during the summer of 1997. Once the afore-mentioned boundary problem is solved using the MAPLE, the solution will be implemented into the existing ADM Code of Pagano and co-workers (1991).

2. Solution Strategy

Starting point of the present research is Pagano's (1991) layerwise axisymmetric solution for fiber-matrix concentric cylinder model, based on Reissner's variational theorem (1950) in conjunction with an equilibrium stress field, in which the radial (r-) dependence is assumed *a priori*. An approximate model was formulated to define the thermoelastic response of a concentric fiber-matrix cylindrical body under axisymmetric boundary conditions. The interfaces between contiguous cylinders may be either continuous or subjected to mixed traction and displacement boundary conditions. The external surfaces may be subjected to mixed boundary conditions that are consistent with the model assumptions but otherwise arbitrary.

In what follows, an improved stress field within a layer is derived starting from Pagano's layerwise axisymmetric fiber-matrix concentric cylinder solution. The stress distribution in the radial direction obtained from the afore-cited layerwise fiber-matrix concentric cylinder model (Pagano, 1991) is improved *a posteriori* through an iterative approach that involves successive substitution of the previously computed strains (or stresses) into the equations of compatibility and equilibrium. A similar procedure was implemented in the post-processing part of a layerwise finite element code for analysis of quasi-three-dimensional laminated plates/shells to obtain a more accurate through-thickness distribution of interlaminar shear stresses (see Chaudhuri and Seide, 1987; and Chaudhuri, 1990). The boundary/interface conditions at $r = \text{constant}$ surfaces are to be satisfied in the pointwise sense, thus eliminating artificial discontinuity in computed $\sigma_\theta(r,z)$ and $\sigma_z(r,z)$ across a layer-interface within the same material.

Pagano's (1991) Model

Pagano (1991) considered an arbitrary region within the body defined by inner and outer radii r_1^k and r_2^k , respectively and end planes $z = z_1, z_2$ as shown in Figure 2. He introduced a right-handed cylindrical coordinate system z, θ, r , and employed a contracted notation in the representation of the stress and strain components, i.e.,

$$\sigma_1 = \sigma_{zz}, \sigma_2 = \sigma_{\theta\theta}, \sigma_3 = \sigma_{rr}, \sigma_5 = \sigma_{rz} \quad (1)$$

and the analogous relation for the engineering strain components ϵ_i ($i=1, 2, 3, 5$). The r, z components of displacement are designated as u, w , respectively.

The form of the stress distribution within the annular region is assumed to be given by

$$\sigma_i = p_{ij} f_j^{(i)} \quad (i = 1, 2, 3, 5; j = 1, 2, \dots, 5) \quad (2)$$

where p_{ij} are functions of z only. In order to avoid confusion with layer superscript k , the index i in $f_j^{(i)}$ is placed in parentheses. For a region in which $r_1 \neq 0$, the functions $f_j^{(i)}$ are defined by

$$\begin{aligned} f_1^{(1)} &= f_1^{(2)} = f_1^{(3)} = \frac{r_2 - r}{r_2 - r_1} \\ f_2^{(1)} &= f_2^{(2)} = f_2^{(3)} = \frac{r - r_1}{r_2 - r_1} \\ f_1^{(5)} &= \frac{r}{r_1} f_1^{(1)}, \quad f_2^{(5)} = \frac{r}{r_2} f_2^{(1)} \quad (r_1 \neq 0) \\ f_3^{(3)} &= r^3 - (r_1^2 + r_1 r_2 + r_2^2)r + r_1 r_2 (r_1 + r_2) \end{aligned} \quad (3)$$

$$f_4^{(3)} = r^2 - (r_1 + r_2)r + r_1 r_2, \quad f_5^{(3)} = \frac{1}{r_1 r_2 r} f_4^{(3)}$$

$$f_3^{(5)} = \frac{(r_1 + r_2)r^2 - (r_1^2 + r_1 r_2 + r_2^2)r}{r_1^2 r_2^2} + \frac{1}{r}$$

with

$$p_{ij} = f_j^{(i)} = 0 \quad (r_1 \neq 0; i = 1, 2 \text{ and } j = 3, 4, 5 \text{ or } i = 5 \text{ and } j = 4, 5) \quad (4)$$

In other words, the functions $f_j^{(i)}$ and the corresponding p_{ij} not displayed in (3) all vanish. In the event that $r_1 = 0$ (fiber core), eqn (3) is replaced by

$$\begin{aligned} f_1^{(1)} &= f_1^{(2)} = f_1^{(3)} = \frac{r_2 - r}{r_2} \\ f_2^{(1)} &= f_2^{(2)} = f_2^{(3)} = f_2^{(5)} = \frac{r}{r_2} \quad (r_1 = 0) \\ f_3^{(3)} &= (r^2 - r_2^2)r, \quad f_4^{(3)} = f_3^{(5)} = (r - r_2)r \end{aligned} \quad (5)$$

with

$$p_{ij} = f_j^{(i)} = 0 \quad (r_1 = 0; i = 1, 2 \text{ and } j = 3, 4 \text{ or } i = 5 \text{ and } j = 1, 4 \text{ or } j = 5) \quad (6)$$

It may be noted that the superscripts k have been omitted in eqns (2) - (6) to avoid unnecessary congestion. The relations (2) - (6) arise by assuming that σ_1 and σ_2 are linear functions of r in the region and then determining the form of the remaining stress components from the equations of equilibrium of axisymmetric elasticity subjected to the following conditions

$$P_{i\alpha}(z) = \sigma_i(r_\alpha, z) \quad (i = 1, 2, 3, 5; \alpha = 1, 2) \quad (7)$$

Thus, the p functions are equal to actual stresses at $r = r_1, r_2$.

The general form of Pagano's (1991) solution for any of the dependent variables $P(z)$ is expressed by

$$P(z) = \sum_i A_i e^{\lambda_i z} + P_p(z) \quad (8)$$

within each constituent where A_i are constants, λ_i are eigenvalues of a determinant, and $P_p(z)$ is a particular solution, which in the present case is a simple polynomial. Further details of the solution procedure including the method for determining the higher order eigenvector and higher order particular solution are discussed by Brown (1992).

Improved Stress Field in a Fiber or Matrix Layer

An examination of Pagano's (1991) axisymmetric fiber-matrix concentric cylinder model reveals that he has assumed a linear variation of axial stress, $\sigma_z(r, z)$ and hoop stress, $\sigma_\theta(r, z)$, with respect to r within a fiber or matrix layer. This is consistent with the assumption of Love-Kirchhoff's thin shell theory. However, he derived the interlaminar shear stress, $\tau_{rz}(r, z)$, and the radial stress, $\sigma_r(r, z)$, using the equations of equilibrium in line with his earlier work (see Pagano, 1969) before substituting these stresses into Reissner's (1950) variational principle. In what follows, improved deformation and stress fields based on successive use of compatibility and equilibrium equations of axisymmetric elasticity theory is derived in the annular and core regions, respectively.

Annular Layer:

Pagano's stress field for an annular (fiber or matrix) layer $k = 1, 2, \dots$ is as shown below:

$$\bar{\sigma}_z^{(k)} = p_{11}^{(k)}(z) \left(\frac{r_2 - r}{r_2 - r_1} \right) + p_{12}^{(k)}(z) \left(\frac{r - r_1}{r_2 - r_1} \right) \quad (9a)$$

$$\bar{\tau}_{rz}^{(k)}(r, z) = p_{51}^{(k)}(z) \frac{r(r_2 - r)}{r_1(r_2 - r_1)} + p_{52}^{(k)}(z) \frac{r(r - r_1)}{r_2(r_2 - r_1)} + p_{53}^{(k)}(z) \left(\frac{(r_1 + r_2)r^2 - (r_1^2 + r_1r_2 + r_2^2)r}{r_1^2 r_2^2} + \frac{1}{r} \right) \quad (9b)$$

$$\begin{aligned} \bar{\sigma}_r^{(k)}(r, z) = & p_{31}^{(k)}(z) \left(\frac{r_2 - r}{r_2 - r_1} \right) + p_{32}^{(k)}(z) \left(\frac{r - r_1}{r_2 - r_1} \right) + p_{33}^{(k)}(z) (r^3 - (r_1^2 + r_1r_2 + r_2^2)r + r_1r_2(r_1 + r_2)) \\ & + p_{34}^{(k)}(z) (r^2 - (r_1 + r_2)r + r_1r_2) + p_{35}^{(k)}(z) (r^2 - (r_1 + r_2)r + r_1r_2) \frac{1}{r_1r_2} \end{aligned} \quad (9c)$$

It is worthwhile to point out here that although the barred stresses satisfy the equations of equilibrium for an axisymmetric elastic body, the corresponding strains fail to satisfy the equations of compatibility in the pointwise sense. It is noteworthy that although both the plane strain and axisymmetric deformations represent two-dimensional states, the latter case requires

four compatibility equations to be satisfied by the strains computed using Pagano's (1991) axisymmetric variational model. This is in contrast to the case of plane strain, wherein only one compatibility equation out of the six is not an identity. However, if the hoop stress, σ_θ , is derived using the exact axisymmetric elasticity based kinematic relations and Hooke's law, then three of the four compatibility equations required by the axisymmetric elasticity theory become identities. The combined kinematic and stress-strain relations for axisymmetric elasticity theory are given as follows (Timoshenko and Goodier, 1959):

$$\tilde{u}(r, z) = r\tilde{\epsilon}_\theta(r, z) = \frac{r}{E_k} [\tilde{\sigma}_\theta(r, z) - \nu_k \{\tilde{\sigma}_r(r, z) + \tilde{\sigma}_z(r, z)\}] \quad (10a)$$

$$\frac{\partial \tilde{u}(r, z)}{\partial r} = \tilde{\epsilon}_r(r, z) = \frac{1}{E_k} [\tilde{\sigma}_r(r, z) - \nu_k \{\tilde{\sigma}_\theta(r, z) + \tilde{\sigma}_z(r, z)\}] \quad (10b)$$

The tilda quantities represent the "improved" stresses, strains and displacements, obtained using the equations of compatibility, equilibrium, etc., but may not satisfy the prescribed $z=\text{constant}$ boundary conditions. Eliminating $\tilde{u}^{(k)}(r, z)$ from eqns (10) and solving for $\tilde{\sigma}_\theta^{(k)}(r, z)$ will yield

$$\tilde{\sigma}_\theta^{(k)}(r, z) = r^{-(1+\nu_k)} \left[\int \left\{ (1+\nu_k) r^{\nu_k} \tilde{\sigma}_r(r, z) + \nu_k r^{(1+\nu_k)} \frac{\partial \tilde{\sigma}_r(r, z)}{\partial r} + \nu_k r^{(1+\nu_k)} \frac{\partial \tilde{\sigma}_z(r, z)}{\partial r} \right\} dr + F_1^{(k)}(z) \right] \quad (11)$$

where $F_1^{(k)}(z)$, $k = 1, 2, \dots$ is a "constant" of integration (with respect to r) and an arbitrary function of z . The hoop stress in an annular layer ($k \neq 0$, may be fiber or matrix) is obtained by substitution of $\tilde{\sigma}_z^{(k)}(r, z)$ and $\tilde{\sigma}_r^{(k)}(r, z)$ from eqns (9a,c) into the right side of eqns (11)

$$\tilde{\sigma}_\theta^{(k)}(r, z) = \sigma_\theta^{(k)*}(r, z) + \frac{1}{r^{1+\nu_k}} F_1^{(k)}(z); \quad k = 1, 2, 3, \dots \quad (12a)$$

where

$$\begin{aligned} \sigma_\theta^{(k)*}(r, z) = & (1+\nu_k) [p_{31}^{(k)}(z) \frac{1}{(r_2 - r_1)} \left(\frac{r_2}{\nu_k + 1} - \frac{r}{\nu_k + 2} \right) + p_{32}^{(k)}(z) \frac{1}{(r_2 - r_1)} \left(\frac{r}{\nu_k + 2} - \frac{r_1}{\nu_k + 1} \right) \\ & + p_{33}^{(k)}(z) \left\{ \frac{r^3}{\nu_k + 4} - (r_1^2 + r_1 r_2 + r_2^2) \frac{r}{\nu_k + 2} + r_1 r_2 (r_1 + r_2) \frac{1}{\nu_k + 1} \right\} + p_{34}^{(k)}(z) \left\{ \frac{r^2}{\nu_k + 3} \right. \\ & \left. - (r_1 + r_2) \frac{r}{\nu_k + 2} + \frac{r_1 r_2}{\nu_k + 1} \right\} + p_{34}^{(k)}(z) \frac{1}{r_1 r_2} \left\{ \frac{r}{\nu_k + 2} - (r_1 + r_2) \frac{1}{\nu_k + 1} + \frac{1}{r \nu_k} r_1 r_2 \right\} \\ & + \nu_k \left[-p_{31}^{(k)}(z) \frac{r}{(r_2 - r_1)(\nu_k + 2)} + p_{32}^{(k)}(z) \frac{r}{(r_2 - r_1)(\nu_k + 2)} + p_{33}^{(k)}(z) \left\{ \frac{3r^3}{(\nu_k + 4)} \right. \right. \\ & \left. \left. - (r_1^2 + r_1 r_2 + r_2^2) \frac{r}{(\nu_k + 2)} \right\} + p_{34}^{(k)}(z) \left\{ \frac{2r^2}{(\nu_k + 3)} - (r_1 + r_2) \frac{r}{(\nu_k + 2)} \right\} \right] \end{aligned}$$

$$+ p_{35}^{(k)}(z) \left\{ \frac{r}{r_1 r_2 (v_k + 2)} - \frac{1}{r v_k} \right\} - p_{11}^{(k)}(z) \frac{r}{(r_2 - r_1)(v_k + 2)} + p_{12}^{(k)}(z) \frac{r}{(r_2 - r_1)(v_k + 2)} \quad (12b)$$

The radial displacement component, $\tilde{u}^{(k)}(r, z)$, in the k th layer can now be obtained by substituting eqns (12) into eqn (10a) as follows:

$$\tilde{u}^{(k)}(r, z) = u^{(k)*}(r, z) + \frac{r^{-v_k}}{E_k} F_1^{(k)}(z); k = 1, 2, \dots \quad (13a)$$

where

$$u^{(k)*}(r, z) = \frac{r}{E_k} \{ \sigma_{\theta}^{(k)*} - v_k (\bar{\sigma}_r^{(k)} + \bar{\sigma}_z^{(k)}) \}; k = 1, 2, \dots \quad (13b)$$

whence $\tilde{\epsilon}_r^{(k)}(r, z)$ can be obtained by direct differentiation with respect to r . Interlaminar shear strain, $\bar{\gamma}_{rz}^{(k)}(r, z)$, can now be obtained using Hooke's law:

$$\bar{\gamma}_{rz}^{(k)}(r, z) = \frac{1}{G_k} \bar{\tau}_{rz}^{(k)}(r, z) \quad (14)$$

It may be noted that the above two strains, $\tilde{\epsilon}_r^{(k)}(r, z)$ (or u) and $\bar{\gamma}_{rz}^{(k)}(r, z)$, are not compatible with the axial strain, $\bar{\epsilon}_z^{(k)}(r, z)$, computed using Hooke's law

$$\bar{\epsilon}_z^{(k)} = \frac{1}{E_k} [\bar{\sigma}_z^{(k)} - v_k (\bar{\sigma}_{\theta}^{(k)} + \bar{\sigma}_r^{(k)})] \quad (15)$$

because the remaining compatibility equation

$$\frac{\partial^2 \epsilon_z}{\partial r^2} + \frac{\partial^2 \epsilon_r}{\partial z^2} = \frac{\partial^2 \gamma_{rz}}{\partial r \partial z} \quad (16a)$$

or the ensuing identity

$$\frac{\partial \epsilon_z}{\partial r} + \frac{\partial^2 u}{\partial z^2} = \frac{\partial \gamma_{rz}}{\partial z} \quad (16b)$$

is not satisfied. It may be noted here that since the radial displacement $\tilde{u}^{(k)}(r, z)$ has already been derived using the appropriate kinematic and constitutive relations of axisymmetric elasticity theory, given by eqns (10), the radial strain, $\tilde{\epsilon}_r^{(k)}(r, z)$, will automatically be compatible with the other two strains of eqn (16a), if $\tilde{u}^{(k)}(r, z)$ is compatible. Hence, the appropriate compatibility equation to be satisfied is eqn (16b) instead of eqn (16a). It is noteworthy that the compatibility equation (16a) is an Euler equation of the principle of minimum complementary energy, while eqn (16b) corresponds to an Euler equation of a yet to be derived variational principle intermediate between the principle of minimum potential energy and Reissner's variational principle.

The axial strain, $\epsilon_z^{(k)}(r,z)$, is therefore, obtained by substituting $\bar{u}^{(k)}(r,z)$ and $\bar{\gamma}_{rz}^{(k)}(r,z)$ given by eqns (13) and (14) into the integrated (with respect to r) version of the compatibility equation (16b) as follows:

$$\bar{\epsilon}_z^{(k)}(r,z) = \epsilon_z^{(k)*}(r,z) + \frac{r^{1-\nu_k}}{E_k} F_1^{(k)}(z)'' + F_2^{(k)}(z); k = 1, 2, \dots \quad (17a)$$

where

$$\epsilon_z^{(k)*}(r,z) = -\int \frac{\partial^2 u^{(k)*}(r,z)}{\partial z^2} dr + \int \frac{\partial \bar{\gamma}_{rz}^{(k)}(r,z)}{\partial z} dr; k = 0, 1, 2, \dots \quad (17b)$$

where $u^{(k)*}(r,z)$ and $\bar{\gamma}_{rz}^{(k)}(r,z)$ are given by eqns (13b) and (14), respectively. The corresponding normal stress, $\bar{\sigma}_z^{(k)}(r,z)$, can be obtained by using Hooke's law as follows:

$$\bar{\sigma}_z^{(k)}(r,z) = \sigma_z^{(k)*}(r,z) + r^{1-\nu_k} F_1^{(k)}(z)'' + E_k F_2^{(k)}(z) + \frac{\nu_k}{r^{1+\nu_k}} F_1^{(k)}(z); k = 1, 2, \dots \quad (18a)$$

where

$$\sigma_z^{(k)*}(r,z) = E_k \epsilon_z^{(k)*}(r,z) + \nu_k \{\bar{\sigma}_r^{(k)}(r,z) + \sigma_\theta^{(k)*}(r,z)\} = 0; k = 1, 2, \dots \quad (18b)$$

It is noteworthy that although the axial strain, $\bar{\epsilon}_z^{(k)}(r,z)$, or its stress counterpart, $\bar{\sigma}_z^{(k)}(r,z)$, satisfies the compatibility eqn (16), it is no longer in equilibrium (in the pointwise sense) with the stresses, $\bar{\tau}_{rz}^{(k)}(r,z)$ and $\bar{\sigma}_r^{(k)}(r,z)$. These stresses are, therefore, rederived from the following two equilibrium equations:

$$\frac{\partial \tau_{rz}}{\partial r} + \frac{\tau_{rz}}{r} + \frac{\partial \sigma_z}{\partial z} = 0 \quad (19a)$$

$$\frac{\partial \sigma_r}{\partial r} + \frac{\sigma_r - \sigma_\theta}{r} + \frac{\partial \tau_{rz}}{\partial z} = 0 \quad (19b)$$

$\bar{\tau}_{rz}^{(k)}(r,z)$ and $\bar{\sigma}_r^{(k)}(r,z)$ can now be obtained from eqns (19) upon integration as follows:

$$\bar{\tau}_{rz}^{(k)}(r,z) = \frac{1}{r} \left[-\int r \frac{\partial \bar{\sigma}_z^{(k)}(r,z)}{\partial z} dr + F_3^{(k)}(z) \right] \quad (20a)$$

$$\bar{\sigma}_r^{(k)}(r,z) = \frac{1}{r} \left[\int \bar{\sigma}_\theta^{(k)} dr - \int r \frac{\partial \bar{\tau}_{rz}^{(k)}(r,z)}{\partial z} dr + F_4^{(k)}(z) \right] \quad (20b)$$

Substitution of $\bar{\sigma}_z^{(k)}(r,z)$, given by eqn (18) into the integrated equilibrium equation (20a) yields

$$\bar{\tau}_{rz}^{(k)}(r,z) = \tau_{rz}^{(k)*}(r,z) - \frac{r^{2-\nu_k}}{3-\nu_k} F_1^{(k)}(z)''' - \frac{\nu_k r^{-\nu_k}}{1-\nu_k} F_1^{(k)}(z)' - E_k \frac{r}{2} F_2^{(k)}(z)' + \frac{F_3^{(k)}(z)}{r} \quad (21a)$$

where

$$\tau_{rz}^{(k)*}(r, z) = -\frac{1}{r} \left[\int r \frac{\partial \sigma_z^{(k)*}(r, z)}{\partial z} dr \right] \quad (21b)$$

Substitution of $\tilde{\sigma}_\theta^{(k)}(r, z)$ and $\tilde{\tau}_{rz}^{(k)}(r, z)$ given by eqns (12) and (21), respectively, into the second integrated equilibrium equation (20b) yields

$$\begin{aligned} \tilde{\sigma}_r^{(k)}(r, z) = & \sigma_r^{(k)*}(r, z) + \frac{r^{3-v_k}}{(3-v_k)(4-v_k)} F_1^{(k)}(z)'''' + \frac{v_k r^{1-v_k}}{(1-v_k)(2-v_k)} F_1^{(k)}(z)'' \\ & - \frac{r^{-1-v_k}}{v_k} F_1^{(k)}(z) + E_k \frac{r^2}{6} F_2^{(k)}(z)'' - F_3(z)' + \frac{F_4(z)}{r} \end{aligned} \quad (22a)$$

where

$$\sigma_r^{(k)*}(r, z) = \frac{1}{r} \left[\int \sigma_\theta^{(k)*}(r, z) dr - \int r \frac{\partial \tau_{rz}^{(k)*}(r, z)}{\partial z} dr \right] \quad (22b)$$

The above procedure loses the end boundary conditions in the axial direction (i.e., at surfaces $z = \text{constant}$), a kind of mathematical Alzheimer's, thus rendering the boundary-value formulation ill-posed. This is due to the fact that the above boundary-value problem has now been transformed into an initial value problem, which is analogous to Hadamard's treatment of the Cauchy problem (see Tikhonov and Arsenin, 1979). This ill-posedness will be removed by deriving appropriate boundary constraint terms, which are functions of r , that help satisfy the end boundary conditions at surfaces $z = \text{constant}$ as follows:

$$\tilde{\epsilon}_z^{(k)}(r, z) = w_{,z}^{(k)}(r, z) = \frac{1}{E_k} \{ \tilde{\sigma}_z^{(k)} - v_k (\tilde{\sigma}_r^{(k)} + \tilde{\sigma}_\theta^{(k)}) \} \quad (23)$$

whence

$$w^{(k)}(r, z) = \frac{1}{E_k} \int \{ \tilde{\sigma}_z^{(k)} - v_k (\tilde{\sigma}_r^{(k)} + \tilde{\sigma}_\theta^{(k)}) \} dz + H_1^{(k)}(r) \quad (24)$$

Substitution of $\tilde{\sigma}_z^{(k)}(r, z)$, $\tilde{\sigma}_r^{(k)}(r, z)$ and $\tilde{\sigma}_\theta^{(k)}(r, z)$, given by eqns (18), (22) and (12), respectively into eqn. (24) gives

$$\begin{aligned} w^{(k)}(r, z) = & \int \epsilon_z^{(k)*}(r, z) dz + \frac{1}{E_k} \int \left\{ r^{1-v_k} F_1^{(k)}(z)'' + E_k F_2^{(k)}(z) + \frac{v_k}{r^{1+v_k}} F_1^{(k)}(z) \right\} dz \\ & - \frac{v_k}{E_k} \int \left\{ \frac{r^{3-v_k}}{(3-v_k)(4-v_k)} F_1^{(k)}(z)'''' + \frac{v_k r^{1-v_k}}{(1-v_k)(2-v_k)} F_1^{(k)}(z)'' - \frac{r^{-1-v_k}}{v_k} F_1^{(k)}(z) \right. \\ & \left. + E_k \frac{r^2}{6} F_2^{(k)}(z)'' - F_3(z)' + \frac{F_4(z)}{r} + \frac{1}{r^{1+v_k}} F_1^{(k)}(z) \right\} dz + H_1^{(k)}(r) \end{aligned} \quad (25a)$$

where

$$\varepsilon_z^{(k)*}(r, z) = \frac{1}{E_k} [\sigma_z^{(k)*}(r, z) - \nu_k \{\sigma_r^{(k)*}(r, z) + \sigma_\theta^{(k)*}(r, z)\}] \quad (25b)$$

Next, using the relationship

$$u_{,z}^{(k)}(r, z) = \frac{1}{G_k} \tilde{\tau}_{rz}^{(k)}(r, z) - w_{,r}^{(k)}(r, z) \quad (26)$$

whence $u^{(k)}(r, z)$ can be obtained as follows:

$$u^{(k)}(r, z) = \frac{1}{G_k} \int [\tilde{\tau}_{rz}^{(k)}(r, z) - w_{,r}^{(k)}(r, z)] dz + H_2^{(k)}(r) \quad (27)$$

The stress, $\sigma_r^{(k)}(r, z)$, can now be obtained using Hooke's law

$$\sigma_r^{(k)}(r, z) = E_k \varepsilon_r^{(k)}(r, z) + \nu_k [\tilde{\sigma}_z^{(k)}(r, z) + \tilde{\sigma}_\theta^{(k)}(r, z)] \quad (28)$$

where

$$\varepsilon_r^{(k)}(r, z) = u_{,r}^{(k)}(r, z) \quad (29)$$

Finally, the stresses $\tau_{rz}(r, z)$ and $\sigma_z(r, z)$ can be derived using the equations of equilibrium (19) as follows:

$$\tau_{rz}(r, z) = \frac{1}{r} \int \left[-\frac{\partial}{\partial r} \{ r \sigma_r(r, z) \} + \tilde{\sigma}_\theta(r, z) \right] dz + H_3^{(k)}(r) \quad (30)$$

$$\sigma_z(r, z) = -\frac{1}{r} \int \left[\frac{\partial}{\partial r} \{ r \tau_{rz}(r, z) \} \right] dz + H_4^{(k)}(r) \quad (31)$$

These expressions are derived using the symbolic language, MAPLE.

Core Region:

Pagano's (1991) stress field for the core layer (core, $k = 0$, always fiber) is as shown below:

$$\bar{\sigma}_z^{(0)}(r, z) = p_{11}^{(0)}(z) \frac{r_2 - r}{r_2} + p_{12}^{(0)}(z) \frac{r}{r_2} \quad (32a)$$

$$\bar{\tau}_{rz}^{(0)}(r, z) = p_{52}^{(0)}(z) \frac{r}{r_2} + p_{53}^{(0)}(z) (r - r_2) r \quad (32b)$$

$$\bar{\sigma}_r^{(0)}(r, z) = p_{31}^{(0)}(z) \left(\frac{r_2 - r}{r_2} \right) + p_{32}^{(0)}(z) \frac{r}{r_2} + p_{33}^{(0)}(z) (r^2 - r_2^2) r + p_{34}^{(0)}(z) (r - r_2) r \quad (32c)$$

The hoop stress, $\bar{\sigma}_\theta^{(0)}(r, z)$, is derived using the same procedure as above:

$$\bar{\sigma}_{\theta}^{(0)}(r,z) = r^{-(1+\nu_0)} \left[\int \left\{ (1+\nu_0)r^{\nu_0} \bar{\sigma}_r(r,z) + \nu_k r^{(1+\nu_0)} \frac{\partial \bar{\sigma}_r(r,z)}{\partial r} + \nu_0 r^{(1+\nu_0)} \frac{\partial \bar{\sigma}_z(r,z)}{\partial r} \right\} dr + F_0^{(0)}(z) \right] \quad (33)$$

where $F_0^{(0)}(z)$, is a "constant" of integration (with respect to r) and an arbitrary function of z . The hoop stress in the core is obtained by substitution of $\bar{\sigma}_z^{(0)}(r,z)$ and $\bar{\sigma}_r^{(0)}(r,z)$ from eqns (32a,c) into the right side of eqn (33)

$$\bar{\sigma}_{\theta}^{(0)}(r,z) = \sigma_{\theta}^{(0)*}(r,z) + \frac{1}{r^{1+\nu_f}} F_0^{(0)}(z) \quad (34a)$$

where

$$\begin{aligned} \sigma_{\theta}^{(0)*}(r,z) = & (1+\nu_0) \left[p_{31}^{(0)}(z) \left(\frac{1}{\nu_0+1} - \frac{r}{r_2(\nu_0+2)} \right) + p_{32}^{(0)}(z) \frac{r}{r_2(\nu_0+2)} + p_{33}^{(0)}(z) \left(\frac{r^3}{\nu_0+4} - \frac{\pi_2^2}{\nu_0+2} \right) \right. \\ & + p_{34}^{(0)}(z) \left(\frac{r^2}{\nu_0+3} - \frac{r_2 r}{\nu_0+2} \right) \left. + \nu_0 \left[-p_{31}^{(0)}(z) \frac{r}{(\nu_0+2)r_2} + p_{32}^{(0)}(z) \frac{r}{r_2(\nu_0+2)} + p_{33}^{(0)}(z) \left(\frac{3r^3}{\nu_0+4} \right. \right. \right. \\ & \left. \left. - \frac{\pi_2^2}{\nu_0+2} \right) + p_{34}^{(0)}(z) \left(\frac{2r^2}{\nu_0+3} - \frac{\pi_2}{\nu_0+2} \right) - p_{11}^{(0)}(z) \frac{r}{r_2(\nu_0+2)} + p_{12}^{(0)}(z) \frac{r}{r_2(\nu_0+2)} \right] \end{aligned} \quad (34b)$$

An examination of eqns (34) reveals that $\bar{\sigma}_{\theta}^{(0)}(r,z)$ becomes unbounded at the core centerline, $r = 0$. Enforcement of the boundedness of $\bar{\sigma}_{\theta}^{(0)}(0,z)$ reduces $F_0^{(0)}(z)$ to 0. However, it may be noted that unlike in the case of an annular region discussed earlier, the radial displacement component, $\bar{u}^{(0)}(r,z)$, will not be obtained using the constitutive relation, given by eqn (10a), because this will result in an over-determined system by assigning one extra condition at $r =$ constant boundary/interface.

Interlaminar shear strain, $\bar{\gamma}_{rz}^{(0)}(r,z)$ and the radial normal strain, $\bar{\epsilon}_r^{(0)}(r,z)$, can now be obtained by substitution of the above stresses into Hooke's law:

$$\bar{\gamma}_{rz}^{(0)}(r,z) = \frac{1}{G_0} \bar{\tau}_{rz}^{(0)}(r,z) \quad (35a)$$

$$\bar{\epsilon}_r^{(0)}(r,z) = \frac{1}{E_0} [\bar{\sigma}_r^{(0)} - \nu_0 (\bar{\sigma}_{\theta}^{(0)} + \bar{\sigma}_z^{(0)})] \quad (35b)$$

It may be noted that the above two strains, $\bar{\epsilon}_r^{(k)}(r,z)$ and $\bar{\gamma}_{rz}^{(k)}(r,z)$, are not compatible with the axial strain, $\bar{\epsilon}_z^{(k)}(r,z)$, computed using Hooke's law

$$\bar{\epsilon}_z^{(0)}(r,z) = \frac{1}{E_0} [\bar{\sigma}_z^{(0)} - \nu_0 (\bar{\sigma}_{\theta}^{(k)} + \bar{\sigma}_r^{(0)})] \quad (35c)$$

because the remaining compatibility equation, given by eqn (16a) is not satisfied. It may be remarked that in the absence of an exact solution to axisymmetric elasticity boundary-value problem, the appropriate variational principle will, as mentioned earlier, demand a different compatibility equation, such as eqn (16a) or (16b).

The axial strain is, therefore, obtained by substituting $\bar{\gamma}_{rz}^{(0)}(r,z)$ and $\bar{\epsilon}_r^{(0)}(r,z)$ given by eqns (35 a,b) into the compatibility equation (16a) as follows:

$$\frac{\partial^2 \bar{\epsilon}_z^{(0)}}{\partial r^2} = -\frac{1}{E_0} \cdot \frac{\partial^2}{\partial z^2} \{ \bar{\sigma}_r^{(0)} - \nu_0 (\bar{\sigma}_\theta^{(0)} + \bar{\sigma}_z^{(0)}) \} + \frac{1}{G_0} \cdot \frac{\partial^2 \bar{\tau}_{rz}^{(0)}}{\partial r \partial z} \quad (36)$$

Further substitution of eqns 32(a,c) and (34) into eqn (36) and integration with respect to r twice lead to

$$\bar{\epsilon}_z^{(0)}(r,z) = \epsilon_z^*(r,z) + rF_1^{(0)}(z) + F_2^{(0)}(z) \quad (37)$$

The corresponding stress, $\bar{\sigma}_z^{(0)}(r,z)$, from

$$\bar{\sigma}_z^{(0)}(r,z) = E_0 \bar{\epsilon}_z^{(0)}(r,z) + \nu_0 \{ \bar{\sigma}_r^{(0)}(r,z) + \bar{\sigma}_\theta^{(0)}(r,z) \} \quad (38)$$

can be expressed in the form:

$$\bar{\sigma}_z^{(0)}(r,z) = \sigma_z^{(0)*}(r,z) + E_0 \{ rF_1^{(0)} + F_2^{(0)} \} \quad (39a)$$

with

$$\sigma_z^{(0)*}(r,z) = E_0 \epsilon_z^{(0)*}(r,z) + \nu_0 \{ \bar{\sigma}_r^{(0)}(r,z) + \bar{\sigma}_\theta^{(0)}(r,z) \} \quad (39b)$$

It may be noted that the stresses $\bar{\sigma}_\theta^{(0)}(r,z)$ and $\bar{\sigma}_z^{(0)}(r,z)$, although compatible, are no longer in equilibrium with the remaining stresses, $\bar{\tau}_{rz}^{(0)}(r,z)$ and $\bar{\sigma}_r^{(0)}(r,z)$, given by eqns (32b,c). These stresses are obtained using the equilibrium equations of axisymmetric elasticity in a manner similar to an annular region, mentioned earlier:

$$\bar{\tau}_{rz}^{(0)}(r,z) = \tau_{rz}^{(0)*}(r,z) - E_0 \left\{ \frac{r^2}{3} F_1^{(0)'}(z) + \frac{r}{2} F_2^{(0)'}(z) \right\} + \frac{F_3^{(0)}}{r} \quad (40a)$$

where

$$\tau_{rz}^{(0)*}(r,z) = -\frac{1}{r} \left[\int r \frac{\partial \sigma_z^{(0)*}(r,z)}{\partial z} dr \right] \quad (40b)$$

and

$$\bar{\sigma}_r^{(0)}(r,z) = \sigma_r^{(0)*}(r,z) + E_0 \left\{ \frac{r^3}{12} F_1^{(0)''}(z) + \frac{r^2}{6} F_2^{(0)''}(z) \right\} - F_3^{(0)'} + \frac{F_4^{(0)}}{r} \quad (41a)$$

where

$$\dot{\sigma}_r^{(0)*}(r,z) = \frac{1}{r} \left[\int \sigma_\theta^{(0)*}(r,z) dr - \int r \frac{\partial \tau_{rz}^{(0)*}(r,z)}{\partial z} dr \right] \quad (41b)$$

Finally, the displacement component, $w^{(0)}(r,z)$ is obtained by utilizing the following constitutive relations:

$$\tilde{\epsilon}_z^{(0)}(r,z) = w_{,z}^{(0)}(r,z) = \frac{1}{E_0} \{ \tilde{\sigma}_z^{(0)} - \nu_0 (\tilde{\sigma}_r^{(0)} + \tilde{\sigma}_\theta^{(0)}) \} \quad (42)$$

The new $\tilde{\epsilon}_z^{(0)}(r,z)$ is now given by

$$\tilde{\epsilon}_z^{(0)}(r,z) = \epsilon_z^{(0)*}(r,z) + r F_1^{(0)}(z) + F_2^{(0)}(z) - \nu_0 \left\{ \frac{r^3}{12} F_1^{(0)''}(z) + \frac{r^2}{6} F_2^{(0)''}(z) \right\} + \frac{\nu_0}{E_0} \left\{ F_3^{(0)'} - \frac{F_4^{(0)}}{r} \right\} \quad (43a)$$

with

$$\epsilon_z^{(0)*}(r,z) = \frac{1}{E_0} [\sigma_z^{(0)*}(r,z) - \nu_0 \{ \sigma_r^{(0)*}(r,z) + \sigma_\theta^{(0)*}(r,z) \}] \quad (43b)$$

$w^{(0)}(r,z)$ can now be obtained by integrating $\tilde{\epsilon}_z^{(0)}(r,z)$ as follows:

$$\begin{aligned} w^{(0)}(r,z) = & w^{(0)*}(r,z) + r \int F_1^{(0)}(z) dz + \int F_2^{(0)}(z) dz - \nu_0 \left\{ \frac{r^3}{12} F_1^{(0)'}(z) + \frac{r^2}{6} F_2^{(0)'}(z) \right\} \\ & + \frac{\nu_0}{E_0} \left\{ F_3^{(0)}(z) - \frac{1}{r} \int F_4^{(0)}(z) dz \right\} + H_0^{(0)}(r) \end{aligned} \quad (44a)$$

with

$$w^{(0)*}(r,z) = \int \epsilon_z^{(0)*}(r,z) dz \quad (44b)$$

The remaining steps for derivation of $u^{(0)}(r,z)$, the stresses $\tau_{rz}(r,z)$ and $\sigma_z(r,z)$ are identical to those for an annular region, and will not be repeated here. These expressions are derived using the symbolic language, MAPLE. Finally, the unknown functions of z and r are currently being determined using MAPLE from boundary/interface conditions in the r - and z -directions, respectively.

This ends the first iteration or cycle. The stresses $\sigma_z^{(k)}(r,z)$ and $\sigma_r^{(k)}(r,z)$ computed above will now be substituted back into eqn (12) to compute the hoop stress $\tilde{\sigma}_\theta^{(k)}(r,z)$ in the second cycle, and the process can be continued for either a prescribed number of cycles, n , or until convergence is reached within certain pre-determined tolerance.

The above procedure introduces $4M$ "constants" of integration, $F_i^{(k)}(z)$, $i = 1, \dots, 4$; $k = 0, 1, 2, \dots, M-1$ — 4 per iteration — which are functions of z , for each iteration. These are determined by using $4M$ appropriate boundary/interface conditions per iteration including those at the axis of symmetry.

Interfacing or Patching of Local Asymptotic Singular and Global Axisymmetric Micromechanical Stress Fields

Zak (1964) was the first to demonstrate interfacing a finite difference based global axisymmetric nonsingular stress field with the singular stress field obtained using a two-dimensional (plane strain) asymptotic solution due to Williams (1952). This approach was illustrated by Zak (1964) by matching the solution of his asymptotic analysis with the numerical finite difference analysis for the free-clamped boundary condition. Zak (1964) arbitrarily chose the point of matching at a distance of 2% of the thickness of the annular region for one of the three stresses, σ_z , σ_r and τ_{rz} , which would yield the stress intensity factor. The remaining two stresses computed using the asymptotic analysis, when multiplied with this factor, would match very closely their global axisymmetric counterparts computed using the finite difference analysis. Since the stress, σ_θ , cannot be derived using the plane strain asymptotic analysis, this stress was left out. Zak (1964) further remarked that if the finite difference based numerical results were available closer to the point of stress singularity, further improvement could be expected.

Pochiraju (1993) in continuing Zak's (1964) research computed the axisymmetric stress field using a highly refined finite element mesh near the point of stress singularity with the help of a commercially available package, ABACUS. He used the standard 8-noded quadrilateral elements with no special singularity formulation. Like Zak (1964), Pochiraju (1993) matched one stress component computed using the FEM (finite elements method) at a point on the fiber matrix interface at a very close distance of the order of 10^{-4} - 10^{-7} times the fiber radius with that obtained from the asymptotic analysis. The scaling or stress intensity factor thus computed when multiplied with the other stress components from the asymptotic analysis or with the same stress component at other angles led to matching of all the stress components from the two analyses at all angles at the same radius. Pochiraju (1993) also established the region of dominance for each term of the asymptotic analysis with an error function.

Although Zak's (1964) and Pochiraju's (1993) approach is effective in the context of two-dimensional finite difference and FEM models, such a brute force model cannot be used in situations such as the present variational axisymmetric model because of the afore-mentioned eigenvalue blow-up or in any three-dimensional modeling environment, where the cost of this degree of refinement will be prohibitive. The proposed research represents a novel attempt to recover a nearly "exact" elasticity solution by successively satisfying the equations of elasticity in the pointwise sense, as opposed to mean-square sense of the finite difference or FEM. The improved stress field can easily be matched with its asymptotic analysis counterpart without having to resort to highly refined layering approach of the existing numerical techniques.

3. Example Problem — Expected Results

As an illustration of the present approach and to examine the fidelity of its predictive capability, the body depicted in Figure 2 will be considered, and the improved results computed using the present compatibility/equilibrium based approach will be compared to the corresponding variational model solution due to Pagano (1991). All the external boundaries are assumed to be traction-free and that the body is subjected to a 1°C temperature rise. As a first step, the fiber is represented as a single solid cylinder or the core, while the matrix is represented as a single annular ring (Pagano's N=1). Both materials are assumed to be isotropic with the following properties

$$\begin{array}{lll} E_f = 413\text{GPa} & \nu_f = 0.2 & \alpha_f = 3.25\mu\text{°C} \\ E_m = 63\text{GPa} & \nu_m = 0.2 & \alpha_m = 3.50\mu\text{°C} \end{array}$$

where the subscripts f and m stand for fiber and matrix, respectively, and the geometric parameters are taken as

$$\frac{l}{a} = 10 \quad \frac{b}{a} = 2$$

The computed results ($M = 2$) will be compared with Pagano's (1991) multi-ring (e. g., $M = 10, 20$, etc.) solution in order to assess the degree of improvement of the computed stress field. Next, the effects of layering and number of iterations ($n > 1$) on the convergence of the present solution will be assessed by way of comparison with the exact nonsingular series solution due to Kurtz and Pagano (1991). Finally, the improved "converged" stress field will be matched with the asymptotic singular stress field at a close distance of the order of 10^{-4} - 10^{-7} times the fiber radius.

4. Closure

A semi-analytical iterative approach for enhancing the existing two-dimensional quasi-continuous axisymmetric stress field for a brittle matrix micro-composite (i. e., a single fiber surrounded by a concentric matrix cylinder), is proposed. In the present approach, the stress distribution in the radial direction obtained from the afore-cited variational model due to Pagano (1991) is improved *a posteriori* through an iterative approach that involves successive substitution of the previously computed strains (or stresses) into the equations of compatibility and equilibrium. A semi-analytical iterative approach for enhancing the existing two-dimensional quasi-continuous axisymmetric stress field for a brittle matrix micro-composite (i. e., a single fiber surrounded by a concentric matrix cylinder), is presented. The existing solution employs Reissner's variational theorem in conjunction with an equilibrium stress field in which the radial (r -) dependence is assumed *a priori*.

In the present approach, the stress distribution in the radial direction obtained from the afore-cited variational model is improved *a posteriori* through an iterative approach that

involves successive substitution of the previously computed strains (or stresses) into the equations of compatibility and equilibrium. The equations of compatibility are selected such that they form Euler equations corresponding to appropriate variational principle, such as the principle of minimum complementary potential energy, etc. The boundary/interface conditions at $r = \text{constant}$ and $z = \text{constant}$ surfaces/interfaces are satisfied in the pointwise sense. The expressions for the improved axisymmetric displacement and stress fields are derived using the symbolic language, MAPLE.

An illustrative thermal stress problem is currently being solved, and will be used to compare with the existing variational solution. When completed, this research will represent a novel semi-analytical post-processing tool to improve solution accuracy and numerical efficiency of existing variational solutions for micro-composites. The final results will be reported in Chaudhuri et al. (to be published).

References

- C. Atkinson, J. Avila, E. Betz, and R. E. Smelser, "The Rod Pull Out Problem, Theory and Experiment," *J. Mech. Phys. Solids*, **30** (1982).
- J. Aveston, G. A. Cooper, and A. Kelly, "Single and Multiple Fracture," *The Properties of Fibre Composites Conference Proceedings*, National Physical Laboratory, pp. 15-26. IPC Science and Technology Press Ltd, Guildford, UK (1971).
- H. W. Brown III, "Analysis of Axisymmetric Micromechanical Concentric Cylinder Model," *Seventeenth Annual Mechanics of Composites Review*, Air Force 86145/09-10-92-150, Dayton, OH (1992).
- B. Budiansky, J. W. Hutchinson, and A. G. Evans, "Matrix Fracture in Fiber-Reinforced Ceramics," *J. Mech. Phys. Solids*, **34**, 167 - 189 (1986).
- R. A. Chaudhuri and P. Seide, "An Approximate Method for Prediction of Transverse Shear Stresses in a Laminated Shell," *Int. J. Solids and Structures*, **23**, 1145 - 1161 (1987).
- R. A. Chaudhuri, "A Semi-Analytical Approach for Prediction of Interlaminar Shear Stresses in Laminated General Shells," *Int. J. Solids and Structures*, **26**, 499 - 510 (1990).
- R. A. Chaudhuri, "A Novel Compatibility/Equilibrium Based Iterative Post-Processing Approach for Axisymmetric Brittle Matrix Composites," Summer Research Report submitted to RDL, October (1996).
- R. A. Chaudhuri, N. J. Pagano, G. P. Tandon and A. H. Khan, "Interfacing of Local Asymptotic Singular and Global Axisymmetric Micromechanical Stress Fields in Brittle Matrix Composites," to be published.
- H. L. Cox, "The Elasticity and Strength of Paper and Other Fibrous Materials," *British J. Appl. Phys.*, **3** (1952).
- Z. Hashin and B. W. Rosen, "The Elastic Moduli of Fiber Reinforced Materials," *J. Appl. Mech.* **31** (1964).

- R. D. Kurtz and N. J. Pagano, "Analysis of the Deformation of a Symmetrically Loaded Fiber Embedded in a Matrix Material," *Engr Composites* , **1** (1991).
- V. K. Luk and L. M. Keer, "Stress Analysis for an Elastic Half Space Containing an Axially-Loaded Rigid Cylindrical Rod," *Int. J. Solids Structures* , **15** (1979).
- L. N. McCartney, "New Theoretical Model of Stress Transfer Between Fibre and Matrix in a Uniaxially Fibre-Reinforced Composite," *Proc. R. Soc. London A* **425**, 215 - 244 (1990).
- N. J. Pagano, "Exact Solution for Composite Laminates in Cylindrical Bending", *J. Compos. Mater.* , **3**, 398 - 411 (1969).
- N. J. Pagano, "Stress Fields in Composite Laminates," *Int. J. Solids Structures*, **14** (1978).
- N. J. Pagano, "Axisymmetric micromechanical stress fields in composites," *Proceedings 1991 IUTAM Symposium on Local Mechanics Concepts for Composite Materials Systems*, p 1. Springer Verlag (1991).
- N. J. Pagano and G. P. Tandon, "Elastic Response of Multi-directional Coated-Fiber Composites," *Comp. Sci. Tech.*, **31** (1988).
- G. Pickett and M. W. Johnson, "Analytical Procedures for Predicting the Mechanical Properties of Fiber Reinforced Composites," *Technical Report AFML-TR-65-220* (1967).
- E. Reissner, "On a Variational Theorem in Elasticity," *J. Math. Phys.*, **29** (1950).
- G. E. Smith and A. J. M. Spencer, "Interfacial Traction in a Fibre-Reinforced Elastic Composite Material," *J. Mech. Phys. Solids* , **18** (1970).
- E. Sternberg, "Load-Transfer and Load-Diffusion in Elastostatics," *Proceedings of the Sixth U.S. National Congress of Applied Mechanics*, The American Society of Mechanical Engineers, New York (1970).
- A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, John Wiley & Sons, New York (1977).
- S. P. Timoshenko and J. N. Goodier, *Theory of Elasticity*, 2nd edn., McGraw-Hill, New York (1959).
- A. R. Zak, "Stresses in the Vicinity of Boundary Discontinuities in Bodies of Revolution," *J. Appl. Mech.*, **31**, 150 - 152 (1964).

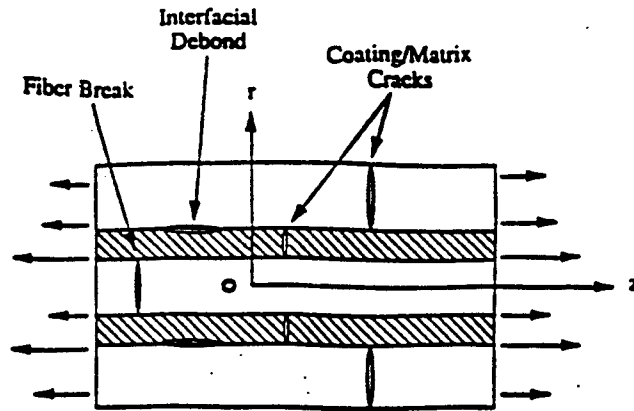


Figure 1. Damage Modes in Brittle Matrix Composites (Pagano, 1991)

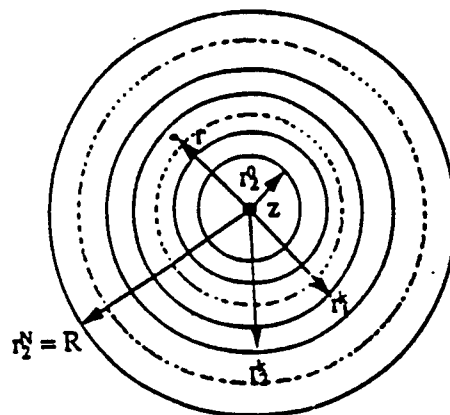


Figure 2. Cross Section of Domain for ADM Analysis (Pagano, 1991)

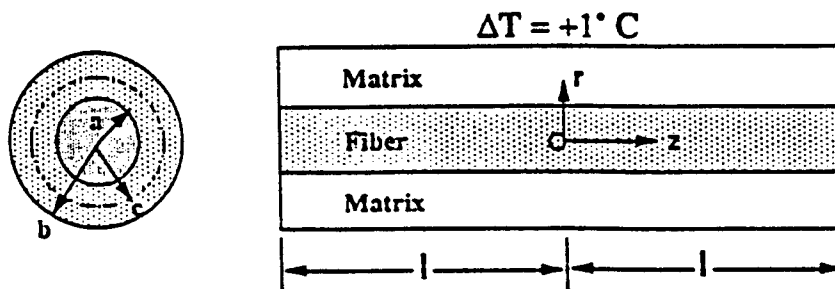


Figure 3. An Illustrative Thermal Stress Problem (Pagano, 1991)

MECHANICAL PROPERTIES AND TEXTURE FORMATION
DURING THE COMPRESSION OF TANTALUM

Robert J. De Angelis
Professor
Department of Mechanical Engineering

University of Nebraska-Lincoln
Lincoln, NE 68588-0656

Final Report for:
AFSOR Summer Research Extension Program
Air Force Research Laboratory
Eglin Air Force Base, FL 32541
AFRL/MNMW

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Air Force Research Laboratory

June 1998

TABLE OF CONTENTS

Abstract	3
Introduction	4
Materials Processing	6
Powder Process	6
Cabot Plate	6
Experimental Methods and Results	6
Powder Process	6
Cabot Plate	7
X-Ray Pole Figures	7
Powder Process	7
Cabot Plate	7
Texture Quantification:	9
Powder Process	9
Cabot Plate	12
Mechanical Properties	13
Powder Process	13
Cabot Plate	14
Discussion	16
References	18
Acknowledgments	18
Figure Captions	19

MECHANICAL PROPERTIES AND TEXTURE FORMATION DURING THE COMPRESSION OF TANTALUM

Robert J. De Angelis

Professor

Department of Mechanical Engineering

University of Nebraska-Lincoln

Abstract

The production of tantalum plate with controlled degrees of anisotropy is essential to obtain reliable performance of explosively formed projectiles (EFP). In this investigation the mechanical response of tantalum was determined in compression for materials produced employing two methods: hot isostatic compression of tantalum powder; and cold forging and rolling of a cast ingot. The fiber texture formation during compression testing was also monitored employing x-ray diffraction techniques. The densities of $\langle 100 \rangle$, $\langle 110 \rangle$ and $\langle 111 \rangle$ directions were determined as a function of compressive strain for the two types of materials. In the hot isostatic compacted tantalum powder specimens the densities of $\langle 111 \rangle$ and $\langle 100 \rangle$ poles were observed to increase very rapidly during the initial 0.7 logarithmic compression strain. Compression of specimens cut in the rolling and transverse directions of the wrought plate showed approximately linear increases of the densities of $\langle 111 \rangle$ and $\langle 100 \rangle$ poles with compressive strain. The rate of increase of the $\langle 111 \rangle$ is about twice the rate of increase of the $\langle 100 \rangle$.

The mechanical response of the specimens in compression were determined. The yield strength of the hot isostatic pressed tantalum was significantly greater than the wrought plate material. Both materials demonstrated yield points at the initiation of plastic deformation. The flow stresses in the powder processed tantalum were four percent greater in specimens with axes of compression in the radial direction than those with axes of compression in the thru thickness direction. In the wrought plate material the specimens with compression axes parallel to the rolling direction had 2.5% lower flow strength than specimens with compression axes parallel to the transverse direction.

MECHANICAL PROPERTIES AND TEXTURE FORMATION DURING THE COMPRESSION OF TANTALUM

Robert J. De Angelis

Introduction:

Processing parameters have a important influence in controlling the defect structure in metals. In the case of pure metals the grain size, grain shape, grain orientation and total defect concentration are controlled by processing. Unquestionably, processing consisting of plastic deformation by compression produces changes in the texture of a polycrystalline material. This change in texture is accompanied by a corresponding change in engineering properties and is the prime cause of anisotropy in polycrystalline metallic materials. Texture or preferred orientation imparts the anisotropic properties of the single crystal to the polycrystalline aggregate. In applications which demand that the material demonstrate extensive plasticity, anisotropic behavior due to the existence of a texture is normally undesirable. This is a important consideration in the processing of face centered cubic metals, such as aluminum and copper. However, it is a critical concern in processing body centered cubic metals, especially tantalum and niobium. To control the anisotropic properties developed in tantalum due to preferred orientation the effects of plastic deformation on texture formation must be precisely determined. To aid in accomplishment of this task requires the availability of a reliable and simple methods to characterize the texture existing at every stage of processing.

Despite the general recognition that detailed texture description is needed to control macroscopic properties very few investigations of materials processing have included the quantification of texture. Since the 1980's the description of material textures or crystal orientations in polycrystalline wires and sheets started to move beyond the crystallographic pole figure representation of texture. The orientation distribution function (ODF) has become the method of choice for presenting the description of material textures. A great advantage of the ODF method of texture representation is that it is capable of giving an accurate representation of the inverse pole figure. In this investigation the inverse pole figure will be employed to quantify

the pole densities of the $\langle 100 \rangle$, $\langle 110 \rangle$ and $\langle 111 \rangle$ poles in tantalum as a function of plastic strain in compression.

The objective of this investigation is to develop the processing parameters that produce tantalum with improved performance in EFP's by controlling the size, shape and crystallographic orientation of the metallic grains. To succeed in this endeavor requires the definition of the relationships existing between the mechanical processing sequence, the crystallographic texture, and the mechanical properties of the material. This approach leads to improved capability of controlling the mechanical properties of tantalum that can be exploited to enhance the design of EFPs.

Materials Processing:

Powder Process : Tantalum powder having the following chemistry: H 28 PPM; O 182 PPM; Si 10 PPM; Fe <5 PPM; Mo 5 PPM; C 16 PPM and N 26 PPM, was consolidated by Ceracon, Inc. (1) The average particle size of the powder was 13 microns and the particle size distribution of the powder was 100/200 mesh - 16.9%, 200/325 mesh -31.3%, and -325 mesh 51.6%. Approximately 2.04 kg of powder was formed into 16.5 mm (0.65 inches) thick and 140 mm (5.5 inches) diameter near net shape preform by cold isostatic pressing of the powder in an elastomer bag at 345 MPa. The preform was hot isostatically forged by Ceracon, Inc. Prior to the final forging the preform was wrapped in tantalum foil placed in a stainless steel container and held for 24 hours at 400°C under a constant vacuum of 10^{-5} torr. The preform was **Ceracon™** forged into a disk 13 mm thick and 150 mm in diameter employing a pressure of 980 MPa, at 1186°C and a dwell time of 30 seconds. The average grain size in the as received plate was determined by Dr. Joel House to be 37.5 μ m.

Cabot Plate: A 305 mm diameter ingot was produced by Cabot Corporation (Cabot ingot number 50014) utilizing a casting process consisting of vacuum arc remelting of a three electron beam processed electrode ingot (Cabot ingot number 80313). The ingot chemistry furnished by Cabot Corp. was: H <5 PPM; O 45 PPM; Si <5 PPM; Fe <5 PPM; Mo <5 PPM; C 15 PPM and N <10 PPM. The cast ingot was turned to 290 mm diameter and side forged at 200°C to a 100 mm thick slab for rolling. The slab was machined clean, annealed at 1050°C and cold rolled to an 8 mm thick plate with the rolling direction perpendicular to the ingot axis. The plate was annealed at 1050°C for one hour. The average grain size determined from nine locations in the plate was 67.1 μ m. (2) The range of grain size observed was between 50.0 and 79.1 μ m with a standard deviation of 7.2 μ m. (2)

Experimental Methods and Results:

The experimental techniques employed in this investigation can be partitioned into three task areas: x-ray pole figure determinations; quantification of pole figure data and mechanical property determination. These tasks were executed at the Air Force Research Laboratory at Eglin Air Force Base, FL (AFRL/MNMW).

Powder Process: Compression specimen, 7.6 mm diameter by 7.6 mm in length, were machined from the **Ceracon™** forged plate. These specimens were produced with the compression axes in

the thru thickness (CZ) and in the radial directions (CR). These specimens were tested in compression at a nominal strain rate of 0.025/s. The compressed specimens were sliced in half transverse to the compression axis, metallographically prepared using standard practice (3) and the prepared specimens were utilized for the determination of x-ray pole figures.

Cabot Plate: Compression specimens, 7.6 mm diameter by 7.6 mm, were machined from the plate which was processed from Cabot ingot 50014. Specimens were produced with compression axes in the rolling (RD) and transverse (TV) directions of the plate. (The rolling direction is normal to the ingot axis and the transverse direction is parallel to the ingot axis.) These specimens were tested in compression at a nominal strain rate of 0.025/s. The compressed specimens were sliced in half transverse to the compression axis, metallographically prepared and employed for x-ray pole figure measurements.

X-Ray Pole Figures: X-ray pole figures were determined using the Siemens D5000 diffractometer at AFRL/MNMW. The specimens were prepared for x-ray analysis and the (110), (200) and (211) pole figures were recorded by Dr. Joel House. All measurements were made on surfaces that had been mid-planes of the compression specimens. These surfaces were exposed by machining. The pole figure data were converted to ODFs (4) and inverse pole figures employing both popLA (5) and Siemens (6) software.

Powder Process: Pole figures were determined on eight specimens, four of the specimen surfaces had normals parallel to the thickness direction of the plate and four of the specimen surfaces had normals parallel to the radial direction of the plate. The specimen orientations along with the mechanical testing parameters for the powder processed specimens are listed in Table I. All eight samples were sectioned to expose the mid-plane of the compressed disk. The mid-plane surface was prepared for x-ray investigation by metallographically polishing and etching (3).

Cabot Plate: Pole figures were determined on eleven specimens, five of the specimen surfaces had normals parallel to the rolling direction of the plate and six of the specimen surfaces had normals parallel to the transverse direction of the plate. The specimen identifications along with the mechanical testing parameters for the Cabot plate specimens are listed in Table I. All eleven samples were sectioned at the mid-plane following compression deformation. The mid-plane surface was prepared for x-ray investigation by metallographically polishing and etching.(3)

Table I : Tantalum Specimens and Mechanical Testing Parameters

Specimen ID	Material	Direction	Stages	L ₀ Inch.	L _F Inch.	Logarithmic strain
CZ-1	Powder	Thru Thick	1	0.297	0.268	-0.100
CZ-2	Powder	Thru Thick	1	0.297	0.228	-0.264
CZ-3	Powder	Thru Thick	2	0.296	0.157	-0.634
CZ-4	Powder	Thru Thick	3	0.296	0.086	-1.236
CZ-5	Powder	Thru Thick	3	0.300	0.141	-0.755
CZ-6	Powder	Thru Thick	1	0.298	0.225	-0.281
CR-2	Powder	Radial	1	0.299	0.225	-0.284
CR-3	Powder	Radial	2	0.300	0.158	-0.641
CR-4	Powder	Radial	3	0.299	0.096	-1.136
CR-5	Powder	Radial	3	0.300	0.141	-0.755
RD-1	Plate 50014	Rolling	1	0.301	0.277	-0.083
RD-2	Plate 50014	Rolling	1	0.300	0.225	-0.288
RD-3	Plate 50014	Rolling	2	0.300	0.153	-0.673
RD-4	Plate 50014	Rolling	3	0.300	0.131	-0.829
RD-5	Plate 50014	Rolling	3	0.300	0.084	-1.273
TV-1	Plate 50014	Transverse	1	0.300	0.258	-0.151
TV-2	Plate 50014	Transverse	1	0.298	0.224	-0.285
TV-3	Plate 50014	Transverse	2	0.300	0.151	-0.686
TV-4	Plate 50014	Transverse	3	0.301	0.130	-0.840
TV-5	Plate 50014	Transverse	3	0.301	0.086	-1.250
TV-6	Plate 50014	Transverse	1	0.300	0.273	-0.094

Texture Quantification:

The texture quantification employed in this investigation is an extension of the method reported by Hosford (7) which is based on inverse pole figures. This method calculates the fraction of poles contained within 2.5° , 7.5° and 12.5° of the $\langle 111 \rangle$, $\langle 100 \rangle$ and $\langle 110 \rangle$ orientations. In this investigation the Hosford analysis was extended to calculate the fraction of non-oriented grains in a specimen and the texture distribution. Three mutually perpendicular orientations of inverse pole figures are presented. One orientation is in the direction of compression axis, the second is in a direction normal to the compression axis, and the third is perpendicular to the initial two directions.

The fraction of non-oriented grains is defined as; $\{1 - \sum (F_{\langle 111 \rangle} + F_{\langle 100 \rangle} + F_{\langle 110 \rangle})\}$, where the $F_{\langle hkl \rangle}$ are the fractions of grains within the chosen angle of the selected $\langle hkl \rangle$ orientation. In the case of an ideally random distribution of grain orientations the $F_{\langle hkl \rangle}$ values for an angle of 12.5° are 0.0965, 0.0715 and 0.1428 for the $\langle 111 \rangle$, $\langle 100 \rangle$ and $\langle 110 \rangle$ orientations respectively. These random values allow the data from any specimen to be expressed as a multiple of the random value ("times random") simply by dividing the measured values by the ideally random values given above.

The texture distribution was also determined by simple extension of these calculations. In the case of an ideally random distribution of grain orientations the $F_{\langle hkl \rangle}$ values for an angle of 7.5° are 0.0351, 0.0290 and 0.0576, and for an angle of 2.5° are 0.0050, 0.0034 and 0.0079 for the $\langle 111 \rangle$, $\langle 100 \rangle$ and $\langle 110 \rangle$ orientations respectively. Data from any specimen can be normalized to times random at the 2.5° , 7.5° and 12.5° angles using the fraction of grains data from the ideal random grain orientation for these three angles. This provides a method to inspect the texture distribution about all of the three major pole orientations; $\langle 111 \rangle$, $\langle 100 \rangle$ and $\langle 110 \rangle$.

Powder Process: The fraction of grains within 12.5° of the $\langle 111 \rangle$, $\langle 100 \rangle$ and $\langle 110 \rangle$ poles in the thickness and radial directions of the CeraconTM forged plate as a function of compressive strain are shown in Fig. 1. During compression of a body centered cubic (BCC) metal the $\langle 111 \rangle$ and $\langle 100 \rangle$ poles align to the compression axis. As these poles align the $\langle 110 \rangle$ poles rotate away from the compression axis. The data in Fig. 1 indicates that the grain rotations occurred during the initial 0.6 compressive strain. At this strain about 80% of the grains are oriented within 12.5° of the $\langle 111 \rangle$ and $\langle 100 \rangle$ poles. At larger strains the values of the fraction of oriented grains

remains constant. The data in Fig. 1 are normalized to random and shown in Fig. 2. These data clearly show the similar behavior of the thickness and radial directions during compression and indicate the $\langle 100 \rangle$ texture (>5 times random) is a little stronger than the $\langle 111 \rangle$ texture (>4 times random).

The data in Fig. 1 were employed to determine the fraction of non-oriented grains as a function of compressive strain as are shown in Fig. 3. In an ideally random sample the fraction of grains outside of 12.5° of the $\langle 111 \rangle$, $\langle 100 \rangle$ and $\langle 110 \rangle$ is 0.689. As shown in Fig. 3 the fraction of non-oriented grains decreases, for both the thickness and radial specimens, from 0.6 at a strain of zero to 0.19 at a strain of 0.6. The texture strengths calculated for a 2.5° angle encircling the $\langle 111 \rangle$, $\langle 100 \rangle$ and $\langle 110 \rangle$ poles for the powder processed specimens are recorded in Table II. The maximum $\langle 111 \rangle$ texture strengths of 21 and 17 were observed in specimens CZ-4 and CR-4.

The texture distribution in the as forged tantalum powder plate is shown in Fig. 4. The maximum strength of the $\langle 111 \rangle$ and $\langle 100 \rangle$ texture components in the thru thickness direction are slightly less than 2.5. The strengths of these components are about equal with the $\langle 100 \rangle$ texture being slightly stronger. Weaker textures exist in the two mutually perpendicular directions normal to the thru thickness direction where the $\langle 110 \rangle$ has a maximum strength of less than 1.5. The texture distributions are broad in all of the three directions of observation. Only the $\langle 100 \rangle$ and $\langle 111 \rangle$ poles show a slight degree of the texture sharpening in the thru thickness direction. The texture distribution in the thru thickness direction sharpens dramatically after being compressed to a strain of 1.236, as shown in Fig. 5. The maximum strengths of the $\langle 111 \rangle$ and $\langle 100 \rangle$ textures are now greater than 20 times random and show a rapid variation with encirclement angle about these two poles. The texture strengths increase from about 5 times random at a deviation angle of 12.5° to 12 times random at a deviation angle of 7.5° and to greater than 20 times random at a deviation angle of 2.5° . As expected this rapid increase in texture with deviation angle in the direction of the compression axis is not observed in the two directions of observation normal to the compression axis.

Table II: Texture Strengths in Units of Times Random for Compression Specimens.

Data Are for 2.5° Encirclement Angle Around the <Hkl> Poles Indicated

Specimen ID	Material	Compression	<111>	<100>	<110>	Nat. Strain
As Forged	Powder	Thru Thick	2.37	2.42	0.5	0
CZ-6	Powder	Thru Thick	7.46	8.08	0.16	-0.281
CZ-3	Powder	Thru Thick	14.7	14.92	0	-0.634
CZ-5	Powder	Thru Thick	9.91	13.24	0	-0.755
CZ-4	Powder	Thru Thick	20.75	20.23	0	-1.236
CZ-2	Powder	Radial	9.07	8.58	0.27	-0.284
CR-3	Powder	Radial	13.53	13.46	0.08	-0.641
CR-5	Powder	Radial	7.96	12.82	0	-0.755
CR-4	Powder	Radial	16.83	13.39	0.1	-1.136
RD-1	Plate 50014	Rolling	0.71	2.1	1.4	-0.083
RD-2	Plate 50014	Rolling	4.27	3.75	0.48	-0.288
RD-3	Plate 50014	Rolling	10.93	5.98	0.42	-0.673
RD-4	Plate 50014	Rolling	10.5	7.49	0.18	-0.829
RD-5	Plate 50014	Rolling	8.72	9.71	0.17	-1.273
TV-6	Plate 50014	Transverse	1.4	2.53	1.15	-0.094
TV-1	Plate 50014	Transverse	0.41	2.51	1.44	-0.151
TV-2	Plate 50014	Transverse	2.98	3.72	0.49	-0.285
TV-3	Plate 50014	Transverse	7.36	6.55	0.04	-0.686
TV-4	Plate 50014	Transverse	7.56	6.72	0.19	-0.84
TV-5	Plate 50014	Transverse	12.22	6.57	0.16	-1.25

The texture distribution in the radial directions sharpen to a lesser degree than the thru thickness direction during compression to a strain of 1.136, as shown in Fig. 6. The maximum strengths of the $\langle 111 \rangle$ and $\langle 100 \rangle$ textures are about 15 times random and show a rapid variation with encirclement angle about these two poles. The texture strengths increase from about 5 times random at a deviation angle of 12.5° to 10 times random at a deviation angle of 7.5° and to 15 times random at a deviation angle of 2.5° . As in the case of the of thru thickness compression, the rapid increase in texture with deviation angle in the direction of the radial compression axis is present to a lesser degree in the two directions of observation normal to the compression axis.

Cabot Plate: The fraction of grains within 12.5° of the $\langle 111 \rangle$, $\langle 100 \rangle$ and $\langle 110 \rangle$ poles in the rolling and transverse directions of the rolled Cabot plate 50014 as a function of compressive strain are shown in Fig. 7. Grain rotation in compression of a BCC metal aligns the $\langle 111 \rangle$ and $\langle 100 \rangle$ poles to the compression axis. As these poles align the $\langle 110 \rangle$ poles rotate away from the compression axis. The data in Fig. 7 indicates that the grain rotations occurred during the initial 0.7 compressive strain. At this strain about 70% of the grains are oriented within 12.5° of the $\langle 111 \rangle$ and $\langle 100 \rangle$ poles. The data contained in Fig. 7 were normalized to random and shown in Fig. 8. Now the different behavior of specimens taken from the rolling and transverse directions during compression is apparent. The texture of the specimens with compression axes in the rolling direction was consistently stronger than the texture in the specimens having compression axes transverse to the rolling direction. The strength of the $\langle 111 \rangle$ texture was generally between 4 and 5 times random which was a little stronger than the $\langle 100 \rangle$ texture (between 3 and 4 times random). The data in Fig. 7 are used to calculate the fraction of non-oriented grains as a function of compressive strain and shown in Fig. 9. In an ideal random sample the fraction of grains outside of 12.5° of the $\langle 111 \rangle$, $\langle 100 \rangle$ and $\langle 110 \rangle$ is 0.689. As seen in Fig. 9 the fraction of non-oriented grains decreases from 0.65, almost random, at a strain of about 0.1 to a value of 0.21 at a strain of 1.2. The data from the rolling and transverse specimens were significantly different. There is a more rapid grain rotation occurring in the specimens having compression axes parallel to the rolling direction, see Fig. 9.

The texture strengths for 2.5° encirclement angle around the $\langle 111 \rangle$, $\langle 100 \rangle$ and $\langle 110 \rangle$ poles for the Cabot plate specimens are also contained in Table II. The maximum $\langle 111 \rangle$ texture strengths of 11 and 12 were observed in specimens RD-3 and TV-5.

The texture distribution in the specimen cut with the compression axis parallel to the rolling direction of the tantalum plate after compression to a strain of 1.257 is shown in Fig. 10. The maximum strength of the texture in the direction of compression is between 10 and 11 times random. The strengths of the $\langle 111 \rangle$ and $\langle 100 \rangle$ textures are equal and their distributions are the same and show no indication of being stronger nearer the $\langle hkl \rangle$ pole. The texture distributions existing in the two perpendicular directions normal to the compression axis are identical. This indicates the formation of a well developed fiber texture. The texture distribution in a specimen cut with the compression axis parallel to the transverse direction of the plate after compression to a strain of 1.250 is shown in Fig. 11. The $\langle 111 \rangle$ texture distribution in the direction of compression is sharper than that observed in the specimen having the compression axis parallel to the rolling direction; see Fig. 10. However, the $\langle 100 \rangle$ texture component is weaker and less sharp than that observed in the specimen having a compression axis parallel to the rolling direction. The texture distributions existing in the two perpendicular directions normal to the compression axis are considerably different in texture strengths. This not only indicates the lack of formation of a uniform fiber texture but may signify the existence of variation in mechanical response in the two perpendicular directions.

Mechanical Properties:

Compression tests at a initial strain rate of 0.025/s were performed using an Instron testing machine having a 50,000 load cell. The mechanical testing parameters are listed in Table I. The compression tests were done in stages, with each stage usually consisting of approximately an engineering strain of 0.3. The number of stages and the total logarithmic strains for each specimen are in Table I. Between each stage of the compression deformation each specimen was remachined to a circular cross-section. The load-displacement data from each stage was transformed to true stress and logarithmic strain. The elastic strain portion of the stress strain curve was subtracted from the data to demonstrate the data as true stress versus logarithmic plastic strain curves.

Powder Process: Typical compression stress versus strain curves for the three stage specimen CZ-4 and CR-4 are shown in Fig. 12 and Fig. 13. These figures contain graphs of engineering stress versus engineering strain, true stress versus logarithmic strain and true stress versus logarithmic plastic strain. The mechanical behavior of these two specimens was almost identical

with both showing a yield point in the second stage of deformation; however only CR-4 showed a yield point in the first stage of compression. The values of the upper yield points, lower yield points and true flow stress at logarithmic strains of 0.1, 0.2 and 0.3 for all powder specimens are listed in Table III. The sample to sample variations are less than ± 10 MPa. However, the specimens cut from the radial direction demonstrate flow stress that are approximately 20 MPa higher than the specimens compressed in the thru thickness direction. The yield stress is about 500 MPa and the flow stress at a strain of 0.3 is 700 MPa. These specimens deformed symmetrically and did not exhibit any eccentricity,

Cabot Plate:

Typical compression stress versus strain curves for the three stage specimen RD-5 and TV-5 are shown in Fig. 14 and Fig. 15 respectively. These figures contain graphs of engineering stress versus engineering strain, true stress versus logarithmic strain and true stress versus logarithmic plastic strain. The mechanical behavior of these two specimens was almost identical with both showing a yield point; however the magnitude of the yield point in specimen TV-5 was greater as can be seen from the data in Table IV. Also included in this table are the values of the upper yield points, lower yield points and true flow stress at logarithmic strains of 0.1, 0.2 and 0.3 for all plate specimens. The sample to sample variations are less than ± 6 MPa; however, the specimens with compression axis in the transverse direction demonstrate flow stress that are approximately 10 MPa higher than the specimens compressed in the rolling direction. The yield stress is about 260 MPa and the flow stress at a strain of 0.3 is 370 MPa. The samples deformed in a non-circular fashion during each stage of compressive deformation. The degree of eccentricity was calculated using, $e = \{ 1 - (b/a)^2 \}^{1/2}$ where b and a are the minimum and maximums in diameter observed. The eccentricities for the specimens are shown in Fig. 16. The specimens tended to show eccentricities of 0.3 except in the last two stages of deformation for RD-5 when the eccentricities were 0.6.

Table III : Yield and Flow Stresses of Tantalum Specimens Produced From Powder.

Specimen ID:	Direction	Upper YP MPa	Lower YP MPa	Flow Stress Strain = 0.1	Flow Stress Strain = 0.2	Flow Stress Strain = 0.3
CZ-1	Thru Thick	495	471	559	NA	NA
CZ-2	Thru Thick	488	473	555	651	661
CZ-3	Thru Thick	504	476	573	657	706
CZ-4	Thru Thick	NA	475	577	663	710
CZ-5	Thru Thick	522	483	581	667	717
CZ-2	Radial	520	484	592	668	716
CR-3	Radial	510	490	588	673	721
CR-4	Radial	495	490	586	671	718
CR-5	Radial	516	489	582	668	716

Table IV : Yield and Flow Stresses of Tantalum Specimens Produced From Cabot Plate 50014

Specimen ID:	Direction	Upper YP MPa	Lower YP MPa	Flow Stress Strain = 0.1	Flow Stress Strain = 0.2	Flow Stress Strain = 0.3
RD-1	Rolling	278	252	297	NA	NA
RD-2	Rolling	262	260	309	342	373
RD-3	Rolling	268	263	310	345	378
RD-4	Rolling	265	249	303	340	371
RD-5	Rolling	256	251	299	334	367
TV-1	Transverse	NA	270	318	NA	NA
TV-2	Transverse	279	266	316	350	380
TV-3	Transverse	261	259	313	351	382
TV-4	Transverse	283	256	307	346	379
TV-5	Transverse	289	256	310	350	384
TV-6	Transverse	284	259	312	NA	NA

Discussion:

Crystallographic texture formation during compression was significantly different in the powder and plate materials. The powdered tantalum formed strong $\langle 111 \rangle$ and $\langle 100 \rangle$ textures at about the same rate with compressive strain, see Fig. 1. In the case of the plate material the $\langle 111 \rangle$ texture formed much faster with compressive strain than did the $\langle 100 \rangle$, see Fig. 7. At high degrees of strain the fraction of grains that are within 12.5° of the $\langle 111 \rangle$ are about 1.5 times the grain fraction that are within 12.5° of the $\langle 100 \rangle$. Based on multiplicity the ratio of $\langle 111 \rangle$ grains to $\langle 100 \rangle$ grains would be 1.33. The result indicates the deformation occurred in an uninhibited fashion. That is there was no inhibition of the cross slip processes and the number grains that rotated to each of the two stable orientation were simply proportional to the number of each type available. The high interstitial content in the powder processed material would be expected to have a significant effect on the cross-slip behavior during deformation. If cross slip is inhibited then $\langle 111 \rangle$ slip on a $\{110\}$ plane has equal probability deforming toward a $\langle 111 \rangle$ or a $\langle 100 \rangle$ pole producing the texture demonstrated by the powder processed material. The $\langle 100 \rangle$ and $\langle 111 \rangle$ grain fractions are equal at high degrees of strain, see Fig. 7.

Chemistry differences between the powder processed and wrought material accounted for large differences in the yield and flow stresses of the tantalum. The average values of the upper yield stress, the lower yield stress and the flow stresses at logarithmic strains of 0.1, 0.2 and 0.3 are given in Table V. The yield stress and flow stress values of the Cabot plate material are about 50% lower than the values observed for the powder processed material. The effects of the processing on the mechanical properties of the tantalum can be seen by comparing the data from the two directions tested for each material. These differences are due to a combination of effects including grain shape, grain morphology and grain orientation or texture. The stress values in the powder processed material are greater in the radial direction by about 15 MPa or about 2%. The Cabot plate material has higher strength and flow stress values in the specimens having a compression axis in the transverse direction. The flow stress values in the transverse specimens are 9 MPa or 2.5% greater than the specimens with their compression axes in the rolling direction. It would be of value to determine the contributions of the various microstructural characteristics to the flow stress; however the data sets are not sufficiently extensive to make the necessary correlations.

Table V: Average of the Yield and Flow Stresses of Tantalum Specimens .

Specimen ID	Direction	Avg. Upper YP MPa	Avg. Lower YP MPa	Flow Stress Strain = 0.1	Flow Stress Strain = 0.2	Flow Stress Strain = 0.3
CZ	Thru Thick	495.6	475.6	569.0	659.5	704.7
CR	Radial	512.6	488.2	587.0	670.0	717.8
CZ + CR	Both	506.6	481.2	577.0	664.7	708.1
RD	Rolling	265.8	255.0	303.6	340.2	372.3
TV	Transverse	279.2	261.0	312.7	349.3	381.3
RD + TV	Both	272.5	258.3	308.5	344.8	376.8

Table VI : Standard Deviations of the Yield and Flow Stresses of Tantalum Specimens

Specimen ID	Direction	Stand. Dev UYP MPa	Stand. Dev LYP MPa	Stand. Dev Strain = 0.1	Stand. Dev Strain = 0.2	Stand. Dev Strain = 0.3
CZ	Thru Thick	8.02	4.50	11.4	7.0	15.4
CR	Radial	10.8	2.90	4.16	2.45	2.36
CZ + CR	Both	12.4	7.60	12.7	7.42	17.6
RD	Rolling	8.13	6.12	5.81	4.64	4.57
TV	Transverse	10.8	5.72	3.98	2.22	2.22
RD + TV	Both	11.4	6.42	6.62	5.87	5.85

The standard deviations in the yield stress and flow stress data is presented in Table VI. These values indicate that the differences observed in the stress values are real. For example the differences between the flow stress in specimens with compression axes parallel and transverse to the rolling direction in the plate material is 9 MPa while the standard deviations for all of the data from these two directions have a maximum of 6.62.

References:

- (1) "Development of Ceracon Forging Process for Producing Near-Net-Shape Tantalum Parts", AFRL-MN-EG-TR-1998-7061, Final Report on Contract F 08630-95-C-0025 December 1997, Henry S. Meeks, Ceracon, Inc., 5300 Claus Road, Riverbank, CA 95367
- (2) "Metallurgical Study of Tantalum Plate for U. S. Air Force", NSWC Memorandum R31:HRL:hrl, 08 MAR 1993, on WL/MNMW Project: "Anisotropic Material Properties of Tantalum", H. Last.
- (3) Ann Marie Kelly, Sherri R. Bingert and Robert D. Reiswig, "New Metallographic Preparation Techniques for Tantalum and Tantalum Alloys", LA-UR-95-2631, Los Alamos National Laboratory.
- (4) H. J. Bunge, Texture Analysis in Materials Science, Butterworth, London 1982.
- (5) U. F. Kocks, J. S. Kallend, H.-R. Wenk, A. D. Rollett and S. I. Wright, "popLA: Preferred Orientation Package - Los Alamos", LA-CC-98-18, Los Alamos National Laboratory, July 1994.
- (6) Siemens Analytical X-ray Instruments, Inc., 6300 Enterprise Lane, Madison, WI 83719.
- (7) William F. Hosford, Final Report for Summer Faculty Research Program, AFOSR, Bolling Air Force Base, DC, 1997.

Acknowledgments:

The data collection, assistance and direction of Dr. Joel House is greatly appreciated. The technical support of Dawn Hattaway, Voncile Ashley, Michael Paul and Bill Peaden at Eglin AFB, AWEF who performed several tasks needed to generate the data included in this report is acknowledged. The assistance of Prof. Peter P. Gillis in reading and commenting on the manuscript is appreciated.

Figure Captions:

Fig. 1 Fraction of Oriented Grains as a Function of Compressive Strain for Powder Processed Tantalum with the Compression Axes in the Thickness and Radial Directions for the Pole Orientations Indicated: ▲ - $\langle 111 \rangle$, ■ - $\langle 100 \rangle$ and ● - $\langle 110 \rangle$.

Fig. 2 The Texture Coefficients as a Function of Compressive Strain for Powder Processed Tantalum with the Compression Axes in the Thickness and Radial Directions for the Pole Orientations Indicated: ▲ - $\langle 111 \rangle$, ■ - $\langle 100 \rangle$ and ● - $\langle 110 \rangle$.

Fig. 3 Fraction of Non -Oriented Grains as a Function of Compressive Strain for Powder Processed Tantalum.

Fig. 4 The Distribution of the Texture Coefficients for the As Forged Powder as a Function of the Angle Encircling the $\langle hkl \rangle$ Poles Indicated. a) In the Direction of Compression, b) In a Direction Normal to the Compression Axis, and c) In a Direction Normal to The Compression Axis and Perpendicular to the Direction in (b).

Fig. 5 The Distribution of the Texture Coefficients for CZ-4, a Thru Thickness Specimen, as a Function of the Angle Encircling the $\langle hkl \rangle$ Poles Indicated. a) In the Direction of Compression, b) In a Direction Normal to the Compression Axis, and c) In a Direction Normal to The Compression Axis and Perpendicular to the Direction in (b).

Fig. 6 The Distribution of the Texture Coefficients for CR-4, a Radial Specimen, as a Function of the Angle Encircling the $\langle hkl \rangle$ Poles Indicated. a) In the Direction of Compression, b) In a Direction Normal to the Compression Axis, and c) In a Direction Normal to The Compression Axis and Perpendicular to the Direction in (b).

Fig. 7 Fraction of Oriented Grains as a Function of Compressive Strain for Cabot Plate Tantalum with the Compression Axes in the Rolling and Transverse Directions for the Pole Orientations Indicated: ▲ - $\langle 111 \rangle$, ■ - $\langle 100 \rangle$ and ● - $\langle 110 \rangle$.

Fig. 8 The Texture Coefficients as a Function of Compressive Strain for Cabot Plate Tantalum with the Compression Axes in the Rolling and Transverse Directions for the Pole Orientations Indicated: ▲ - $\langle 111 \rangle$, ■ - $\langle 100 \rangle$ and ● - $\langle 110 \rangle$.

Fig. 9 Fraction of Non -Oriented Grains as a Function of Compressive Strain for Cabot Plate Tantalum for the Specimen Orientations Indicated.

Fig. 10 The Distribution of the Texture Coefficients for RD-5, a Rolling Direction Specimen, as a Function of the Angle Encircling the $\langle hkl \rangle$ Poles Indicated. a) In the Direction of Compression, b) In a Direction Normal to the Compression Axis, and c) In a Direction Normal to The Compression Axis and Perpendicular to the Direction in (b).

Fig. 11 The Distribution of the Texture Coefficients for TV-5, a Transverse Specimen, as a Function of the Angle Encircling the $\langle hkl \rangle$ Poles Indicated. a) In the Direction of Compression, b) In a Direction Normal to the Compression Axis, and c) In a Direction Normal to The Compression Axis and Perpendicular to the Direction in (b).

Fig. 12 The Compression Stress Versus Strain Curves for Thru Thickness Specimen CZ-4. a) Engineering Stress - Engineering Strain, b) True Stress - Logarithmic Strain, and c) True stress - Logarithmic Plastic Strain.

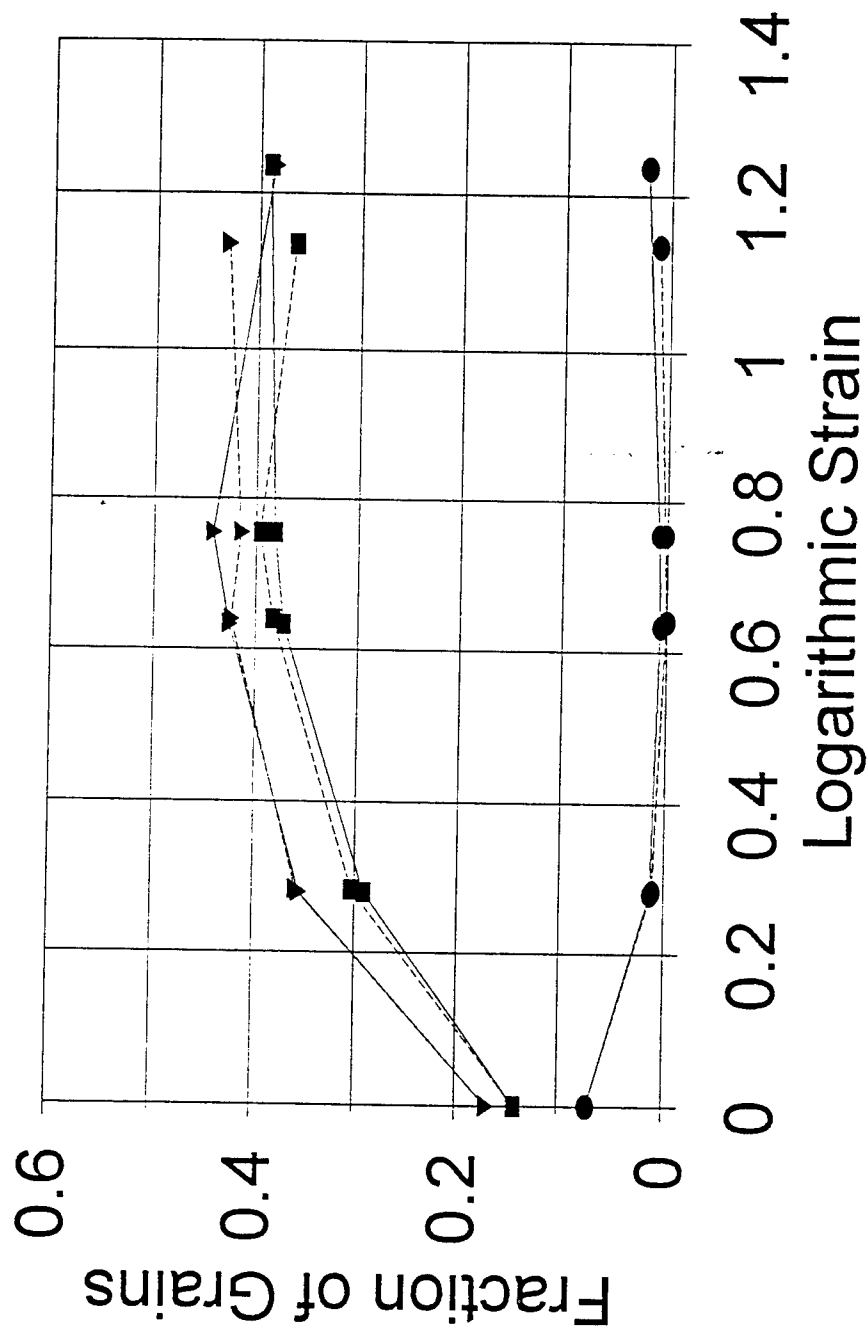
Fig. 13 The Compression Stress Versus Strain Curves for Radial Specimen CR-4. a) Engineering Stress - Engineering Strain, b) True Stress - Logarithmic Strain, and c) True stress - Logarithmic Plastic Strain.

Fig. 14 The Compression Stress Versus Strain Curves for Rolling Direction Specimen RD-5. a) Engineering Stress - Engineering Strain, b) True Stress - Logarithmic Strain, and c) True stress - Logarithmic Plastic Strain.

Fig. 15 The Compression Stress Versus Strain Curves for Transverse Direction Specimen TV-5. a) Engineering Stress - Engineering Strain, b) True Stress - Logarithmic Strain, and c) True stress - Logarithmic Plastic Strain.

Fig. 16 Eccentricity at Each Stage of Deformation of Specimens Taken from the Cabot Plate: (a) With Compression Axes Oriented in the Rolling Direction and (b) With Compression Axes Oriented in the Transverse Direction.

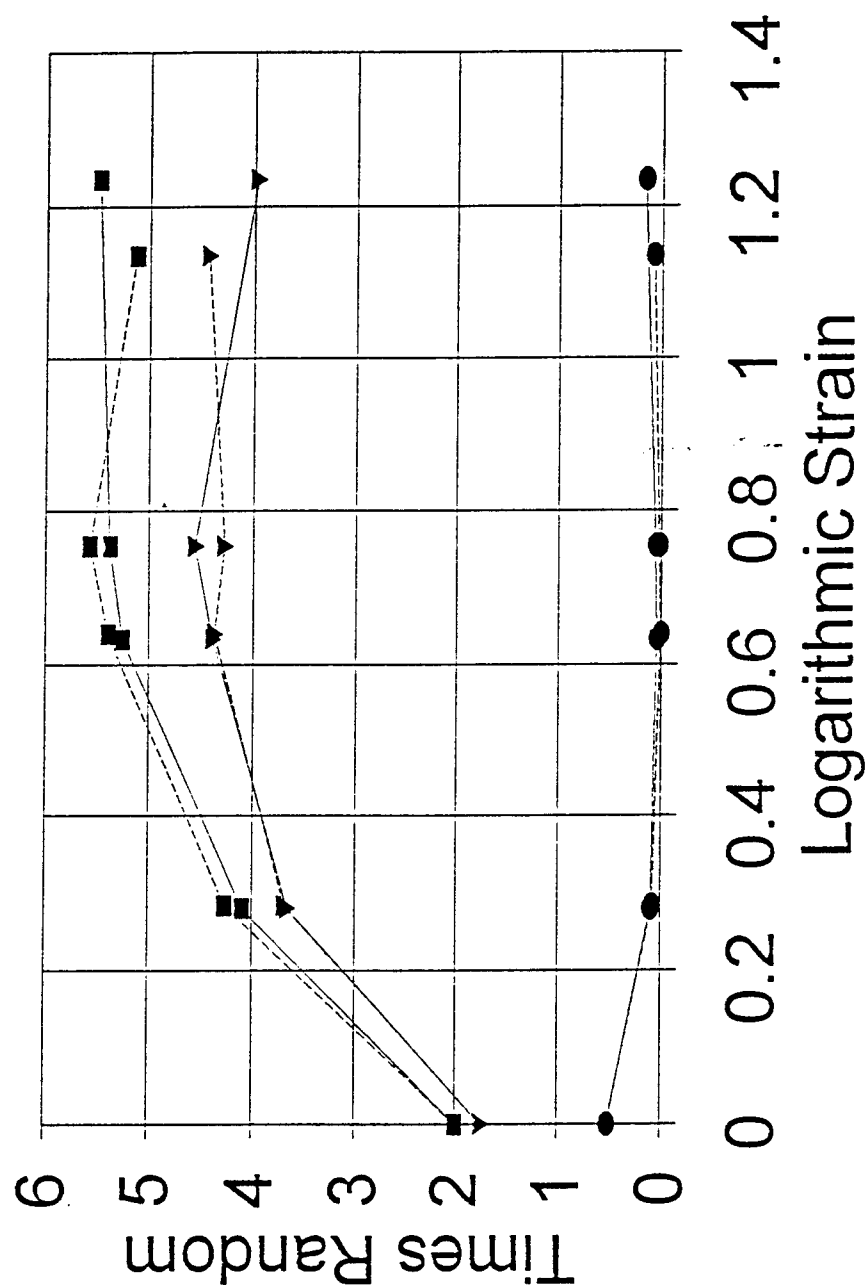
Fraction of Oriented Grains Powdered Tantalum



Thickness Direction Radial Direction

Fig. 1 Fraction of Oriented Grains as a Function of Compressive Strain for Powder Processed Tantalum with the Compression Axes in the Thickness and Radial Directions for the Pole Orientations Indicated: ▲ - $\langle 111 \rangle$, ■ - $\langle 100 \rangle$ and ● - $\langle 110 \rangle$.

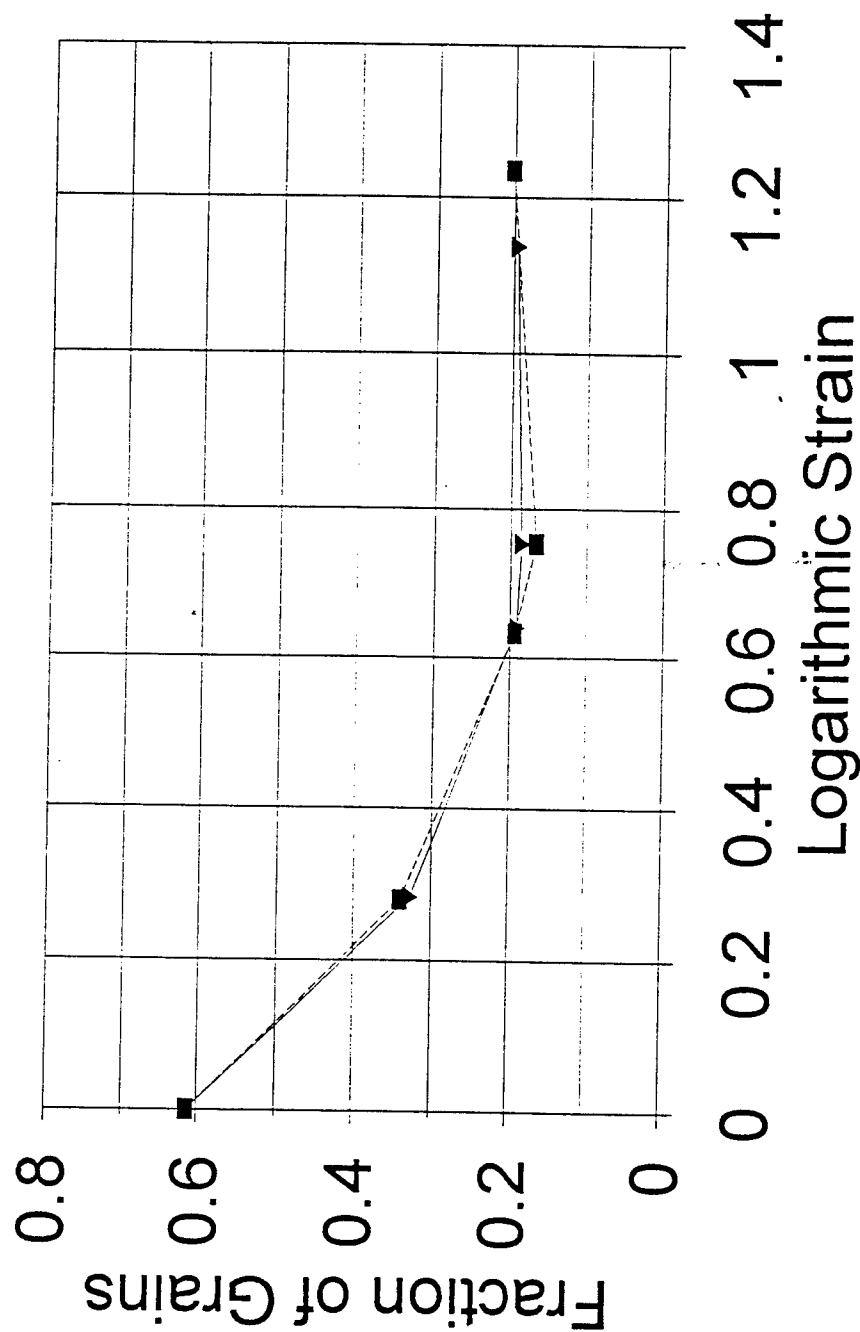
Texture Coefficients Powdered Tantalum



Thickness Direction Radial Direction

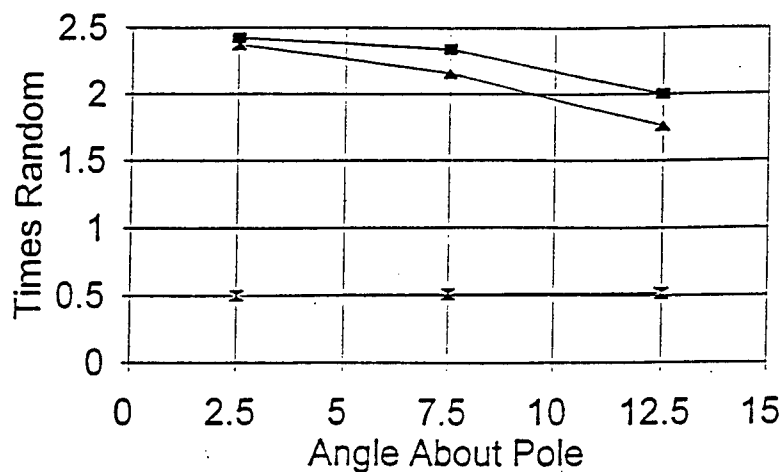
Fig. 2 The Texture Coefficients as a Function of Compressive Strain for Powder Processed Tantalum with the Compression Axes in the Thickness and Radial Directions for the Pole Orientations Indicated: \blacktriangle - $\langle 111 \rangle$, \blacksquare - $\langle 100 \rangle$ and \bullet - $\langle 110 \rangle$.

Fraction of Non-Oriented Grains Powdered Tantalum

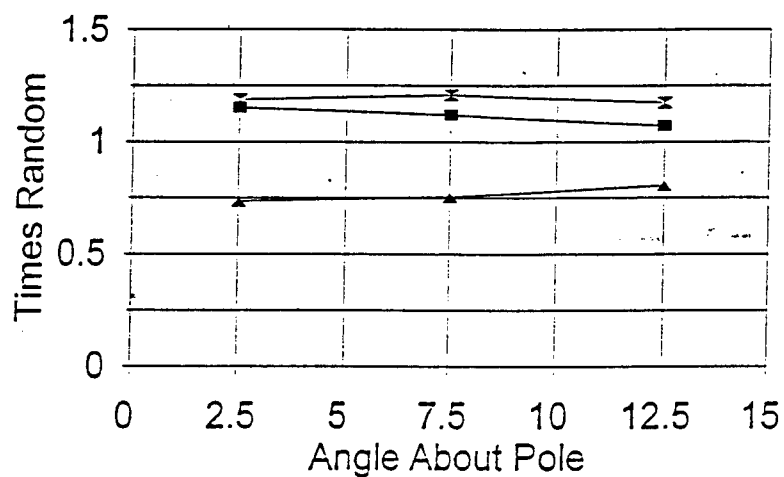


■ Thickness Direction ▴ Radial Direction

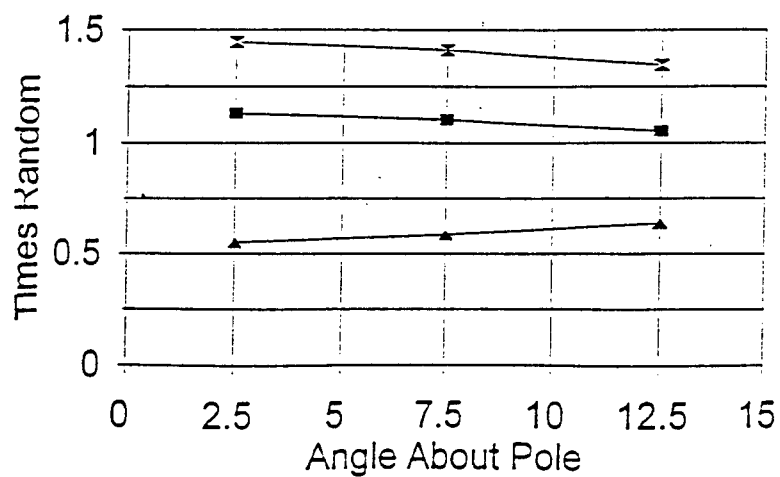
Fig. 3 Fraction of Non -Oriented Grains as a Function of Compressive Strain for Powder Processed Tantalum.



(a)



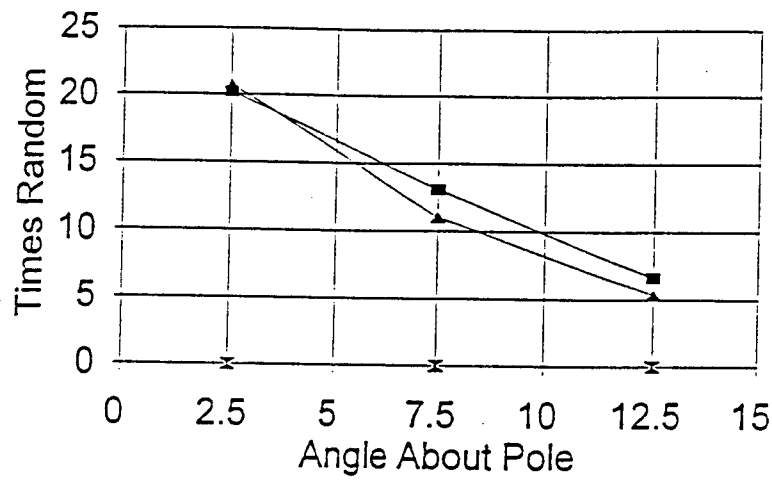
(b)



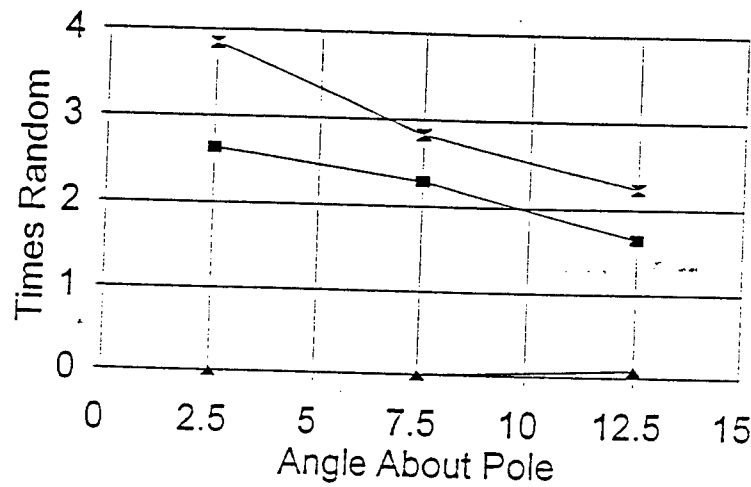
(c)

—▲— <111> —■— <100> —x— <110>

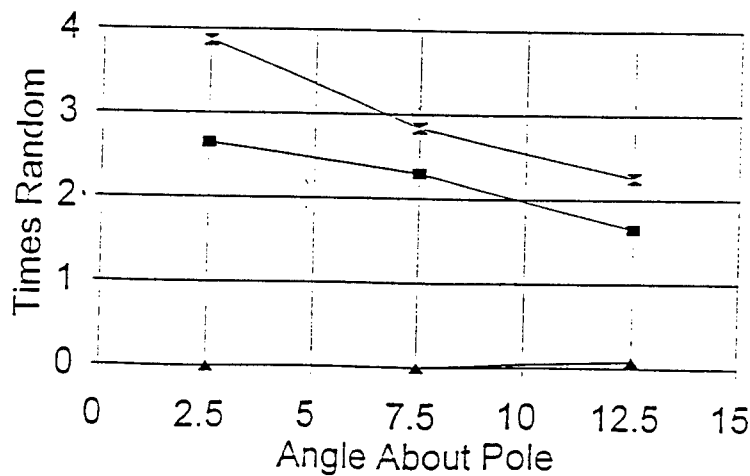
Fig. 4 The Distribution of the Texture Coefficients for the As Forged Powder as a Function of the Angle Encircling the $\langle hkl \rangle$ Poles Indicated. a) In the Direction of Compression, b) In a Direction Normal to the Compression Axis, and c) In a Direction Normal to The Compression Axis and Perpendicular to the Direction in (b).



(a)



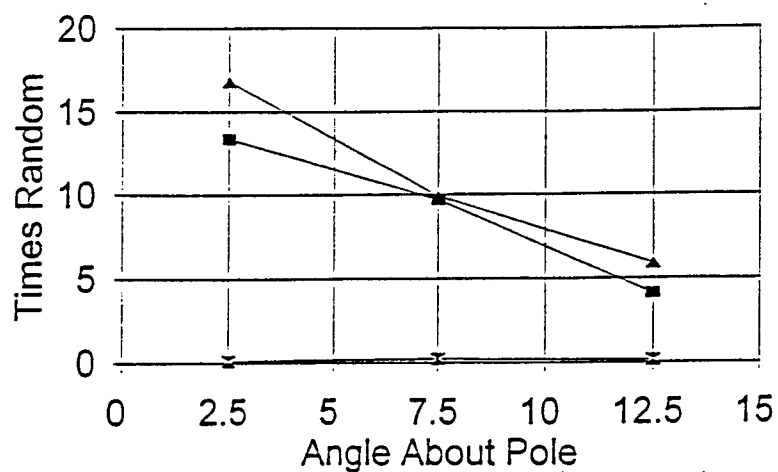
(b)



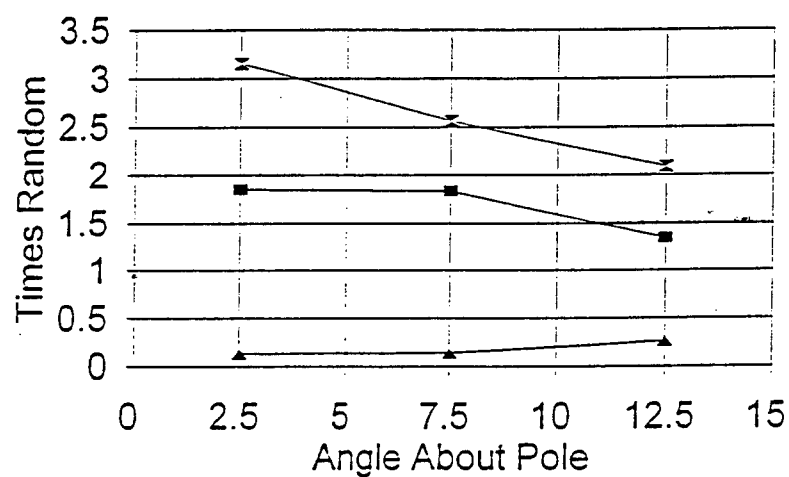
(c)

—▲— <111> —■— <100> —x— <110>

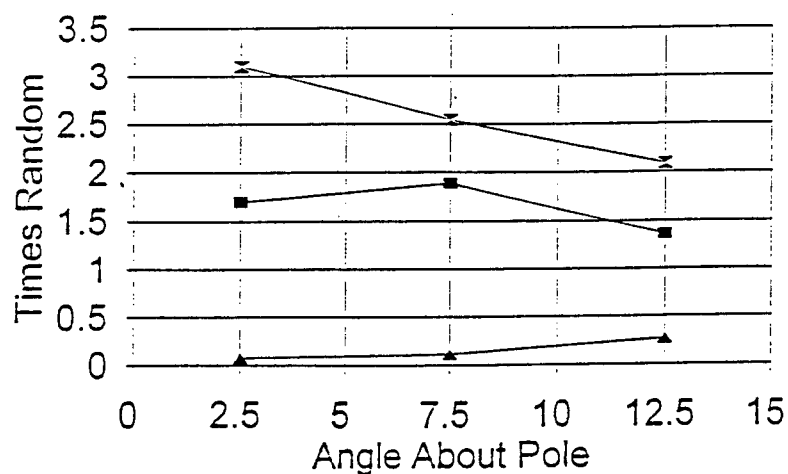
Fig. 5 The Distribution of the Texture Coefficients for CZ-4, a Thru Thickness Specimen, as a Function of the Angle Encircling the $\langle hkl \rangle$ Poles Indicated. a) In the Direction of Compression, b) In a Direction Normal to the Compression Axis, and c) In a Direction Normal to The Compression Axis and Perpendicular to the Direction in (b).



(a)



(b)

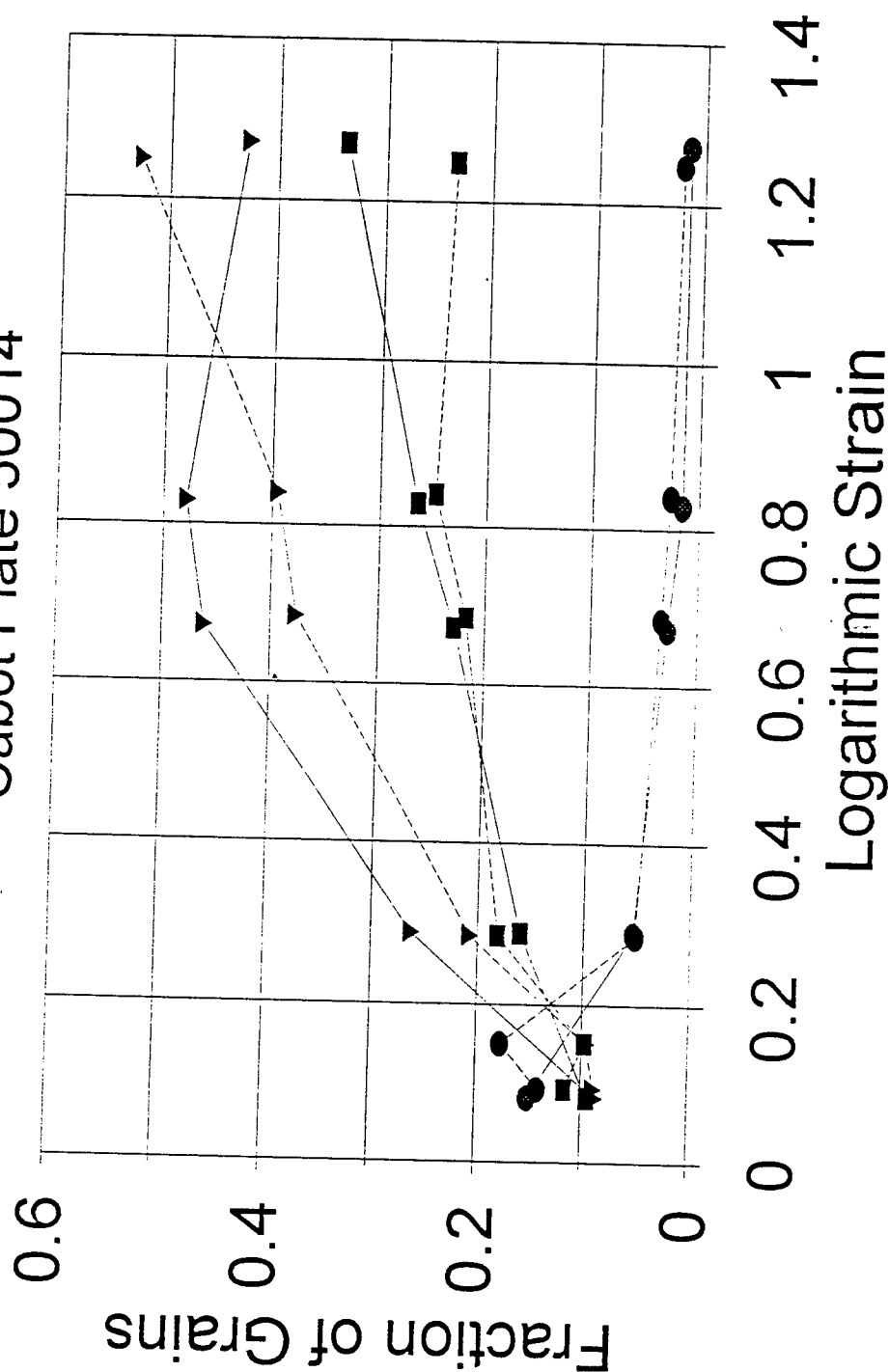


(c)

—▲— <111> —■— <100> —x— <110>

Fig. 6 The Distribution of the Texture Coefficients for CR-4, a Radial Specimen, as a Function of the Angle Encircling the $\langle hkl \rangle$ Poles Indicated. a) In the Direction of Compression, b) In a Direction Normal to the Compression Axis, and c) In a Direction Normal to The Compression Axis and Perpendicular to the Direction in (b).

Fraction of Oriented Grains Cabot Plate 50014



— Rolling Direction - - - Transverse Direction

Fig. 7 Fraction of Oriented Grains as a Function of Compressive Strain for Cabot Plate Tantalum with the Compression Axes in the Rolling and Transverse Directions for the Pole Orientations Indicated: \blacktriangle - $\langle 111 \rangle$, \blacksquare - $\langle 100 \rangle$ and \bullet - $\langle 110 \rangle$.

Texture Coefficients Cabot Plate 50014

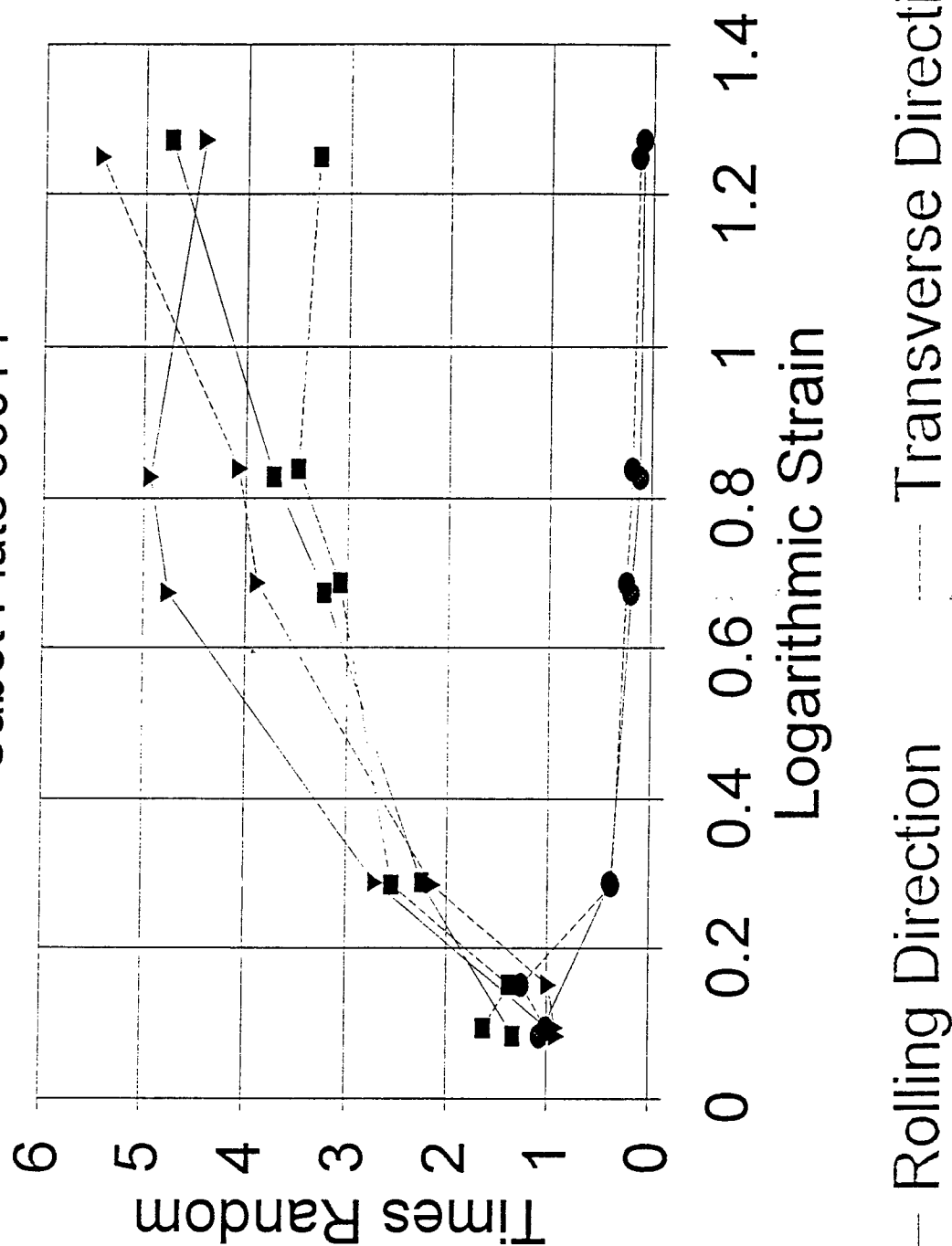


Fig. 8 The Texture Coefficients as a Function of Compressive Strain for Cabot Plate Tantalum with the Compression Axes in the Rolling and Transverse Directions for the Pole Orientations Indicated: \blacktriangle - $\langle 111 \rangle$, \blacksquare - $\langle 100 \rangle$ and \bullet - $\langle 110 \rangle$.

Fraction of Non-Oriented Grains Cabot Plate 50014

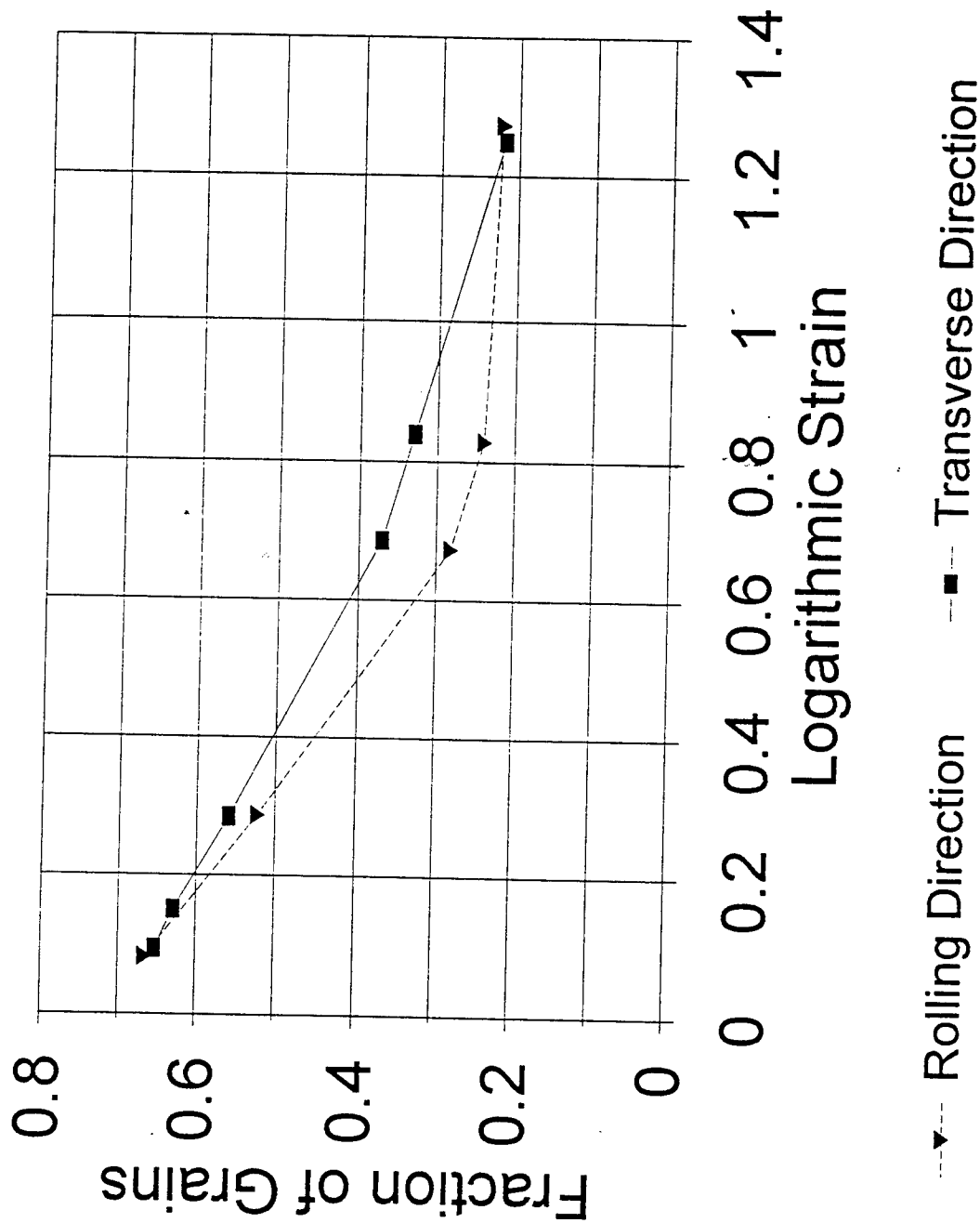
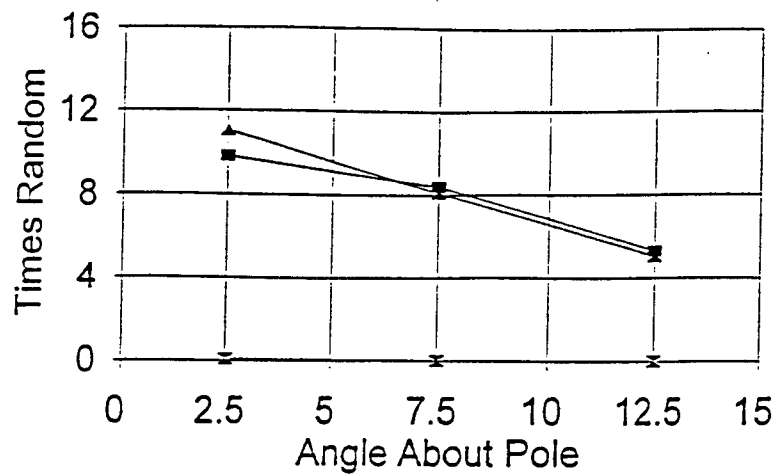
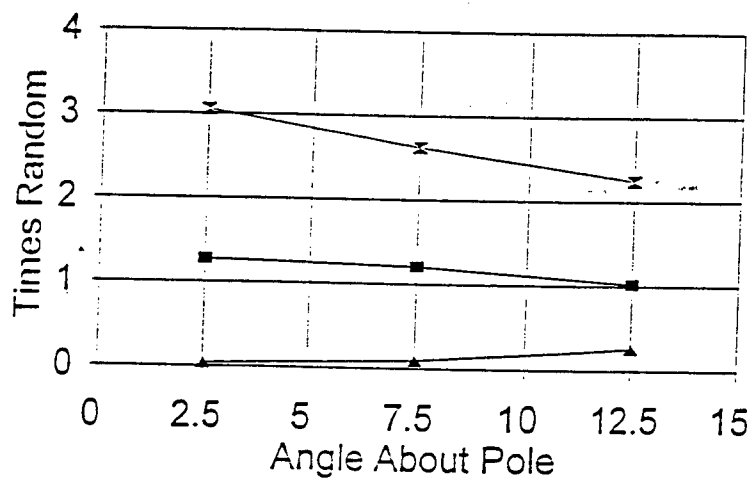


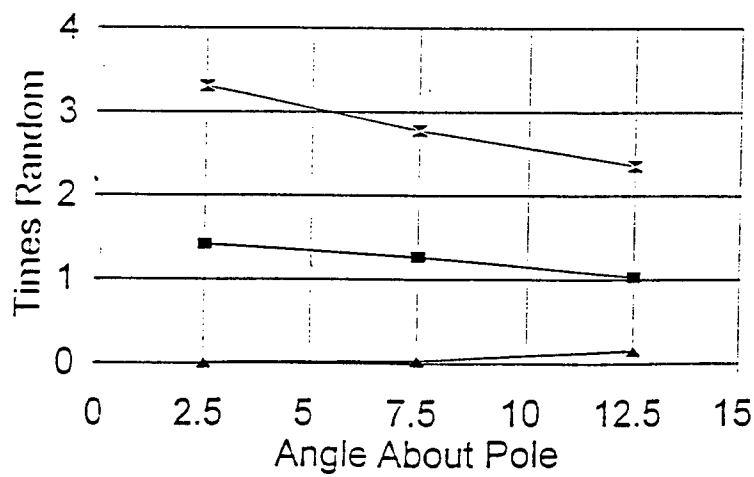
Fig. 9 Fraction of Non -Oriented Grains as a Function of Compressive Strain for Cabot Plate Tantalum for the Specimen Orientations Indicated.



(a)



(b)



(c)

—△— <111> —■— <100> —×— <110>

Fig. 10 The Distribution of the Texture Coefficients for RD-5, a Rolling Direction Specimen, as a Function of the Angle Encircling the $\langle hkl \rangle$ Poles Indicated. a) In the Direction of Compression, b) In a Direction Normal to the Compression Axis, and c) In a Direction Normal to The Compression Axis and Perpendicular to the Direction in (b).

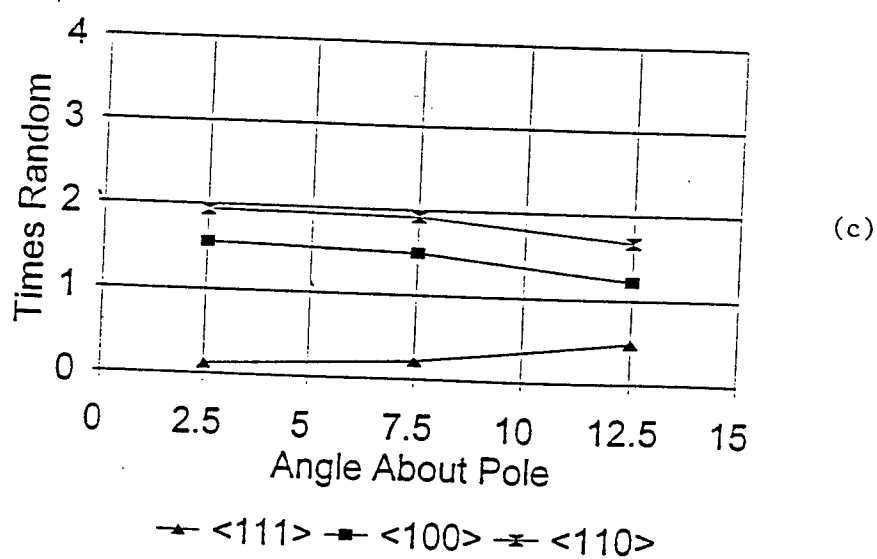
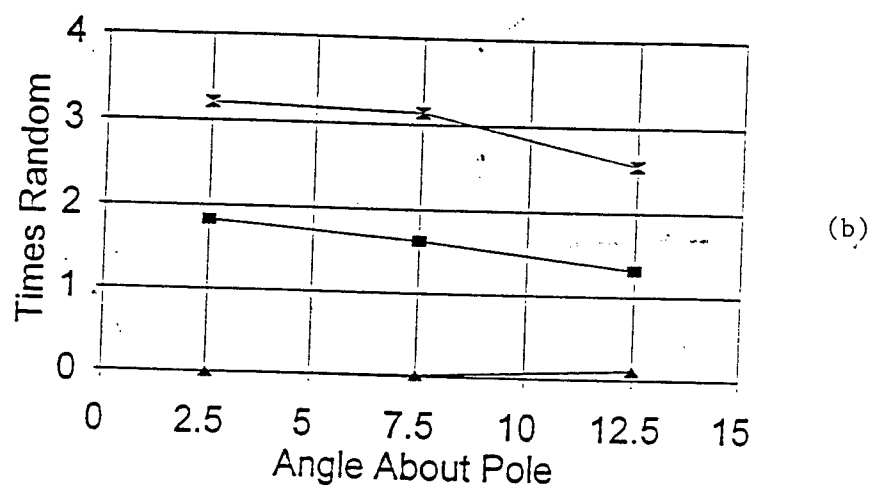
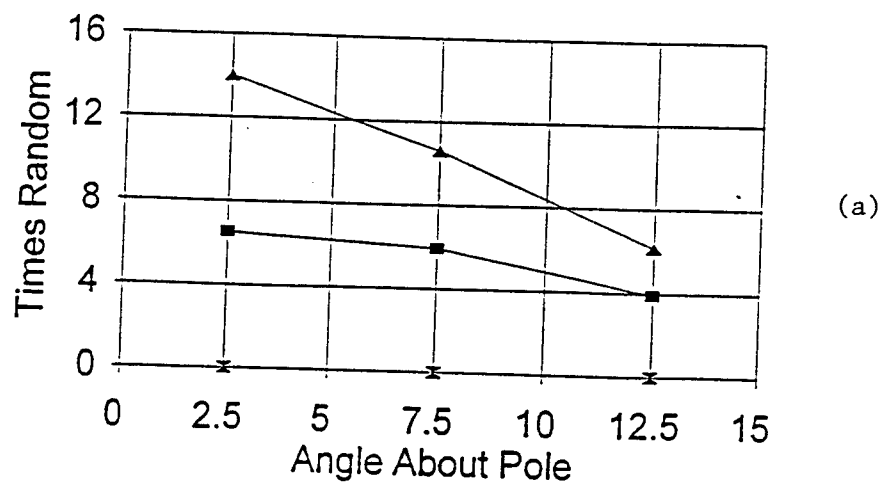
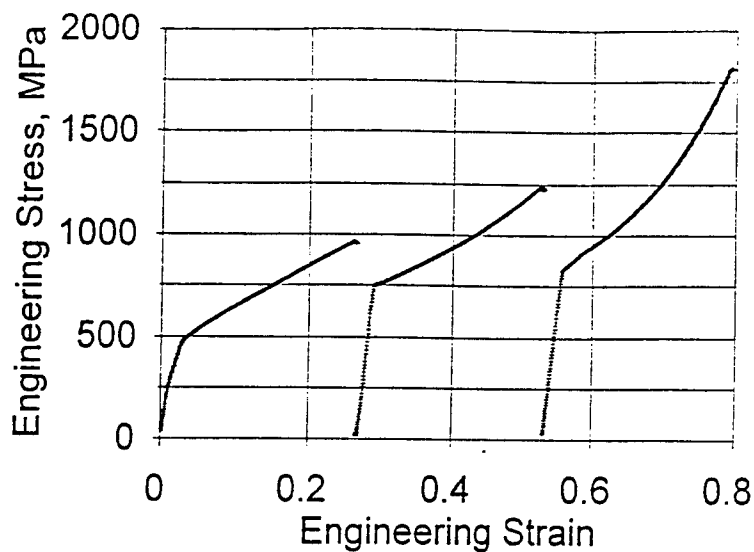
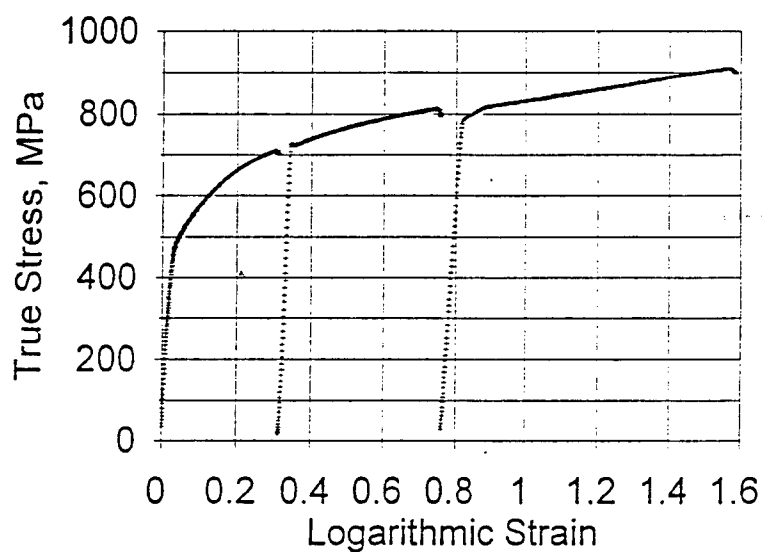


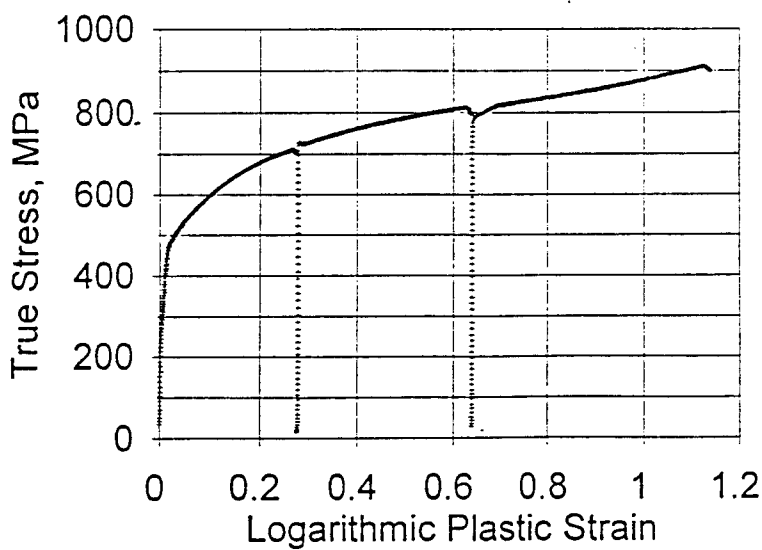
Fig. 11 The Distribution of the Texture Coefficients for TV-5, a Transverse Specimen, as a Function of the Angle Encircling the $\langle hkl \rangle$ Poles Indicated. a) In the Direction of Compression, b) In a Direction Normal to the Compression Axis, and c) In a Direction Normal to The Compression Axis and Perpendicular to the Direction in (b).



(a)

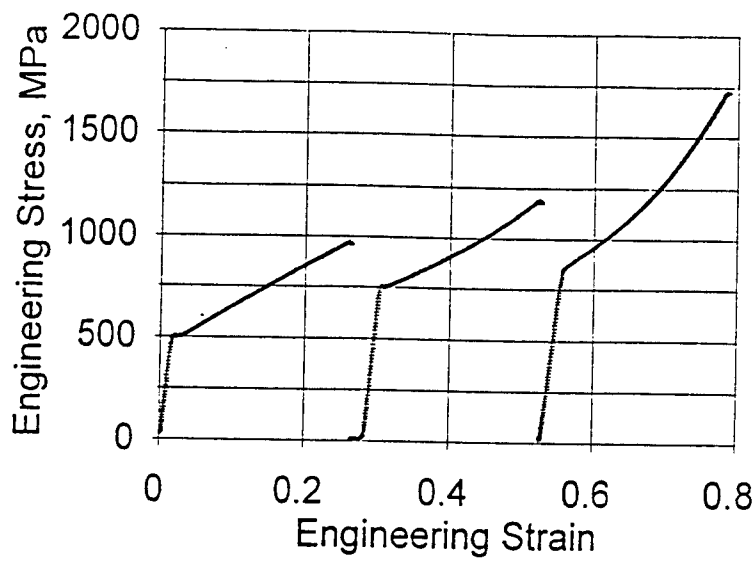


(b)

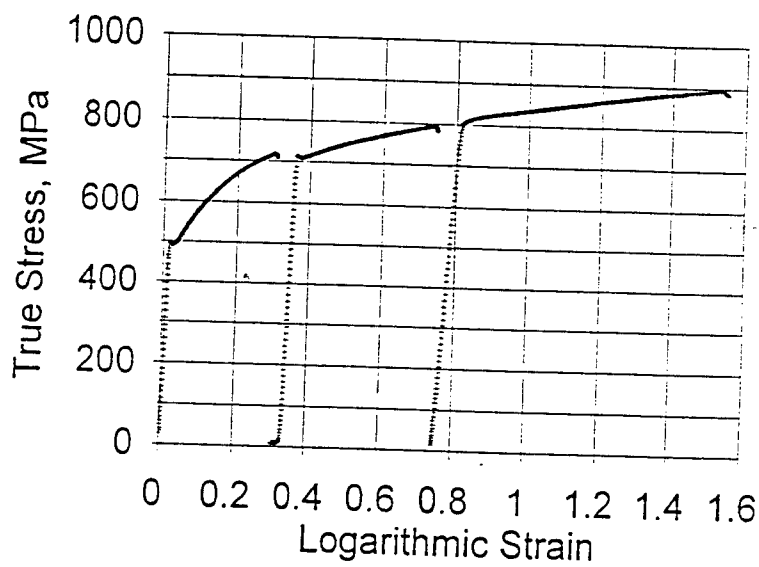


(c)

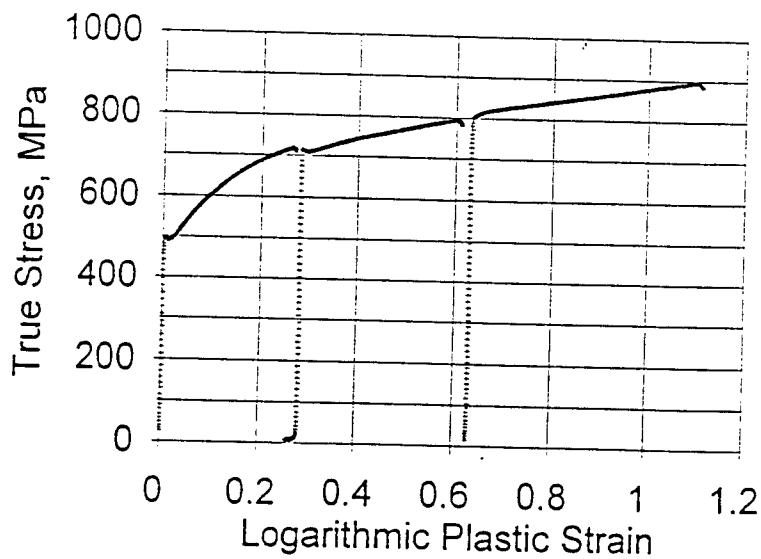
Fig. 12 The Compression Stress Versus Strain Curves for Thru Thickness Specimen CZ-4. a) Engineering Stress - Engineering Strain, b) True Stress - Logarithmic Strain, and c) True stress - Logarithmic Plastic Strain.



(a)



(b)



(c)

Fig. 13 The Compression Stress Versus Strain Curves for Radial Specimen CR-4. a) Engineering Stress - Engineering Strain, b) True Stress - Logarithmic Strain, and c) True stress - Logarithmic Plastic Strain.

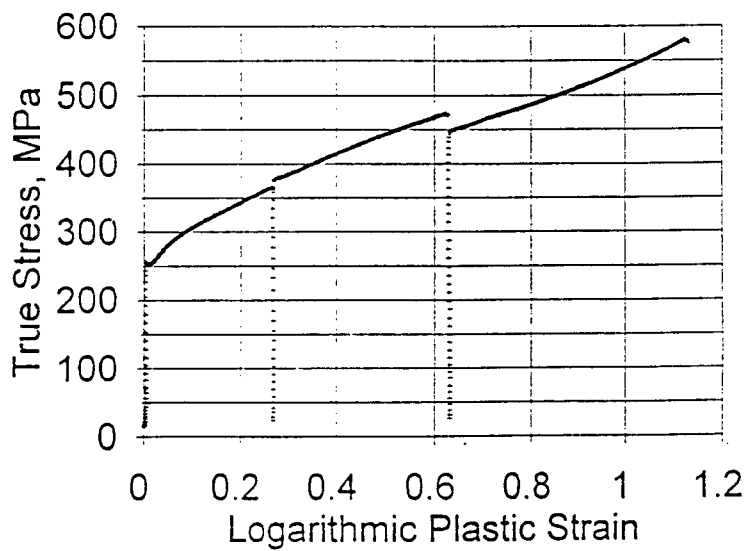
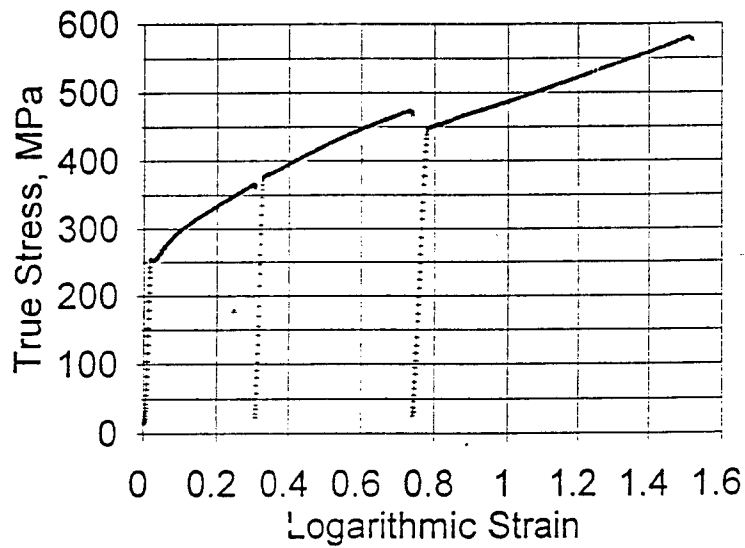
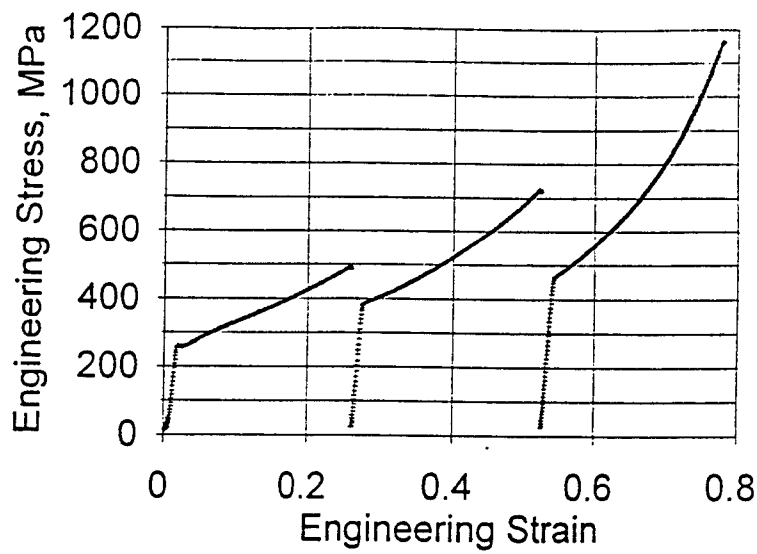
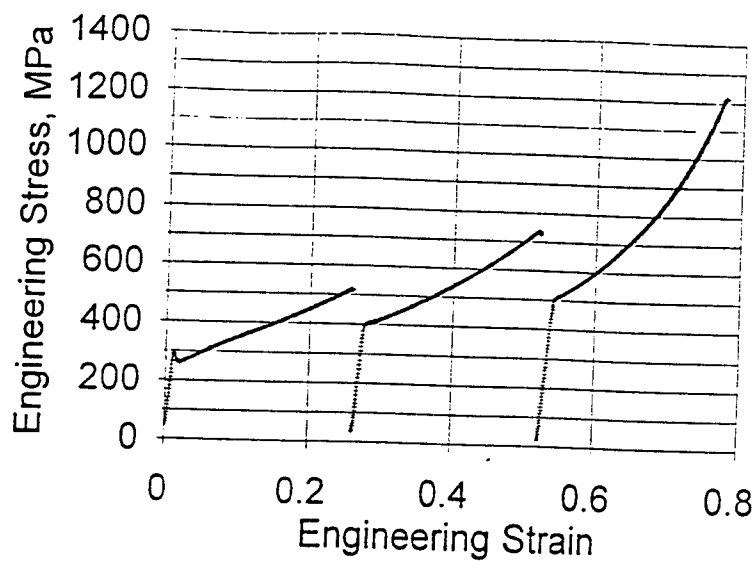
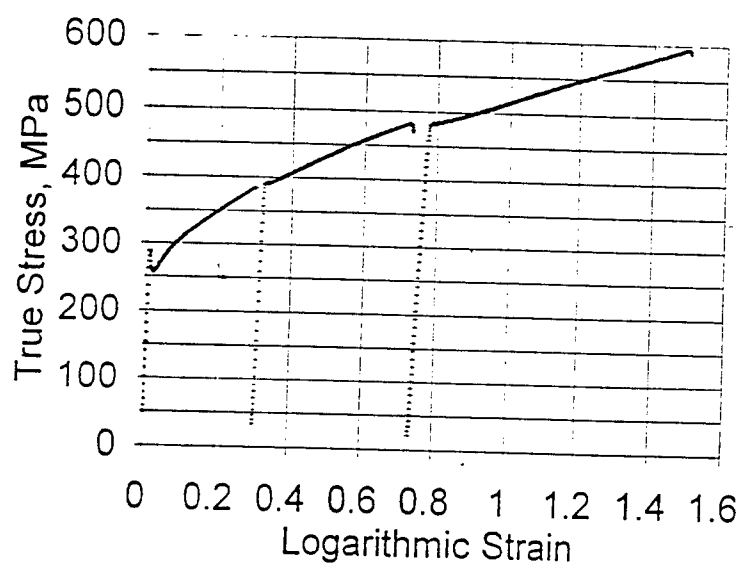


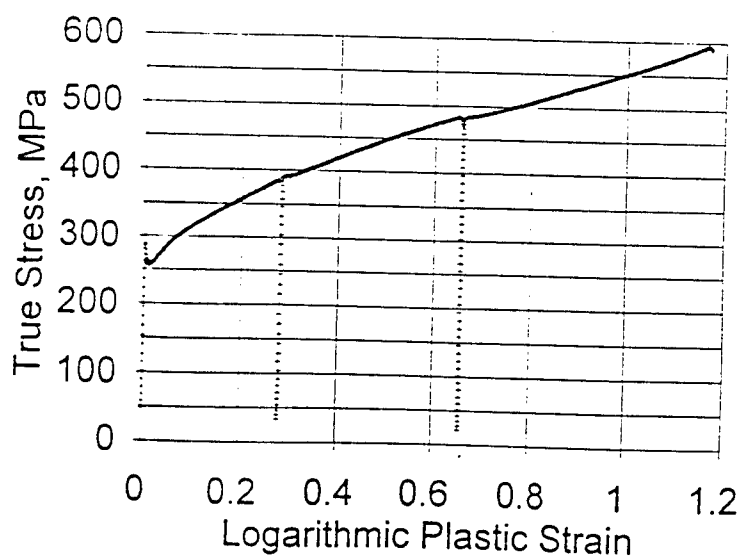
Fig. 14 The Compression Stress Versus Strain Curves for Rolling Direction Specimen RD-5. a) Engineering Stress - Engineering Strain, b) True Stress - Logarithmic Strain, and c) True stress - Logarithmic Plastic Strain.



(a)



(b)



(c)

Fig. 15 The Compression Stress Versus Strain Curves for Transverse Direction Specimen TV-5. a) Engineering Stress - Engineering Strain, b) True Stress - Logarithmic Strain, and c) True stress - Logarithmic Plastic Strain.

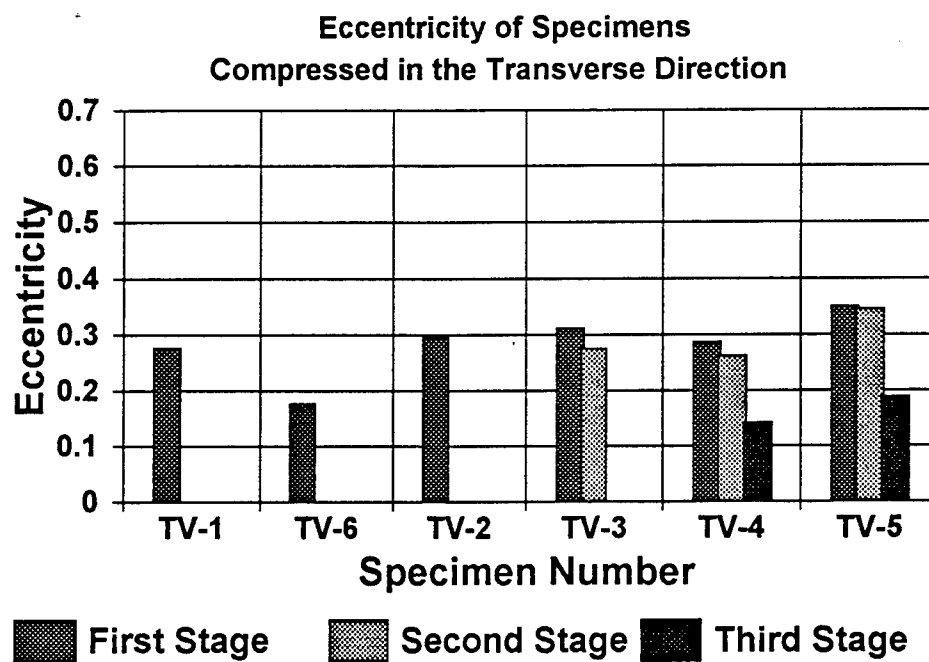
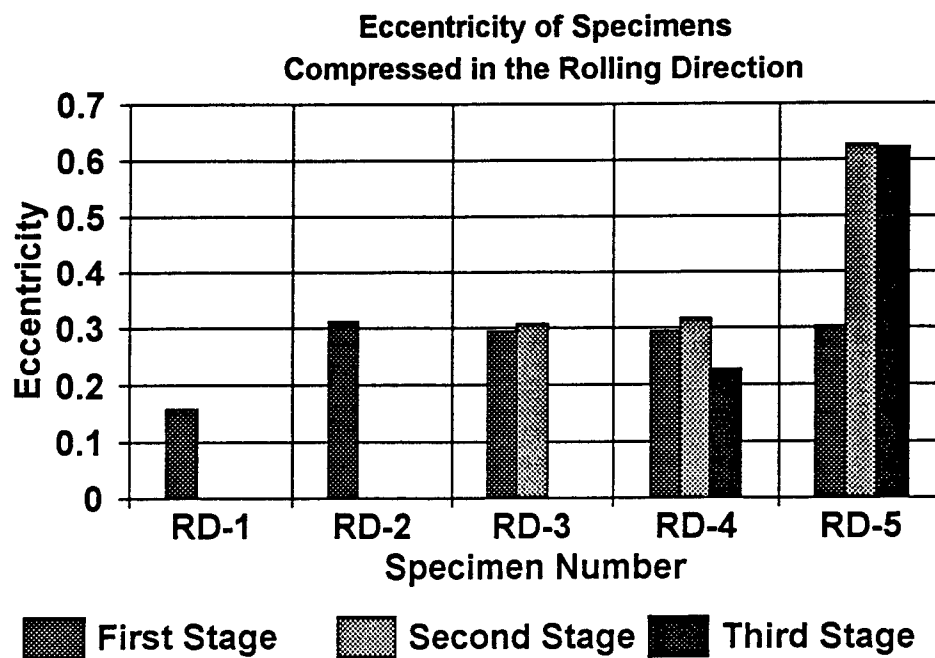


Fig. 16 Eccentricity at Each Stage of Deformation of Specimens Taken from the Cabot Plate: (a) With Compression Axes Oriented in the Rolling Direction and (b) With Compression Axes Oriented in the Transverse Direction.

THE STUDY OF A TRANSVERSE JET IN A SUPERSONIC
CROSS-FLOW USING ADVANCED DIAGNOSTICS

Gregory S. Elliott
Assistant Professor
Department of Mechanical and Aerospace Engineering

Rutgers University
98 Brett Road
Piscataway, NJ 08854-0909

Final Report for:
Summer Research Extension Program
Wright Laboratory

Sponsored by:
Air force Office of Scientific Research
Bolling Air force Base, DC

and
Wright Laboratory

December 1997

THE STUDY OF A TRANSVERSE JET IN A SUPERSONIC CROSS-FLOW USING ADVANCED DIAGNOSTICS

Gregory S. Elliott
Assistant Professor
Department of Mechanical and Aerospace Engineering
Rutgers University

Abstract

Transverse injection of multi-hole sonic injectors into a Mach 3 flow has been experimentally investigated using optimization techniques. The technique allows for the optimization of the positions of multiple injectors based on the efficiency of the mixing as defined by the mixing area, penetration depth, and growth rate of the jet at three locations downstream from the injector center. These quantities were obtained by passing a spanwise sheet of light from the second harmonic of a Nd:YAG pulsed laser through the flow which illuminated the jet mixing seeded with ethanol. The optimization was obtained using Simulated Annealing with Iterative Improvement which was able to improve the mixing compared to the single jet with the same total exit area. Optimized geometries were found when "best" mixing was defined by the mixing area and penetration depth improving the base line case by 82% and 8% respectively for the two injector geometry. For the three injector geometry, the improvement was 66% for mixing area, 2% for penetration, and 12 % for growthrate. Also, a figure of merit combining all mixing definitions was defined and optimized showing an improvement of 43% and 9% for the two and three injector geometries, respectively. Further research beyond the scope of this preliminary investigation will involve studying other optimization routines applied to experimental analysis and experiments which would explain the improvements in mixing given above.

THE STUDY OF A TRANSVERSE JET IN A SUPERSONIC CROSS-FLOW USING ADVANCED DIAGNOSTICS

Gregory S. Elliott

Introduction

Compressible mixing layers have been the subject of much interest motivated by applications such as supersonic air-breathing propulsion systems and reduced heat signatures of high-speed aircraft. However, supersonic mixing is not as effective as their subsonic counterparts to compressibility effects. The transverse jet in a supersonic cross flow (illustrated in Figure 1) has been studied in the past, motivated by its application in several proposed supersonic propulsion systems, in which the qualitative flow features such as the three-dimensional bow shock, separation shock, jet shock structure, and the upstream and downstream separation zones were observed. Beyond this, recent experiments have been conducted to quantify the mixing and penetration of the jet [Lee et al., 1992; Hermanson and Winter, 1993; and Gruber et al., 1997]. Schetz and Billig [1966] found that the penetration could be parameterized by the jet-to-free stream momentum flux, given by:

$$J = \frac{(\gamma PM^2)_j}{(\gamma PM^2)_e} \quad (1)$$

where γ is the specific heat ratio, P is the static pressure, M is the Mach number, and the subscripts j and e indicate the jet and free stream quantities, respectively. A variety of geometries have been investigated to enhance the mixing characteristics of the transverse jet, including swept ramps [Hartfield et al., 1994 and Donohue et al., 1994], multiple ports [Cox et al., 1994 and Fuller et al., 1996], inclined injection [McCann and Bowersox, 1996], and circular and non-circular jet

geometries [Gruber et al., 1995]. It was observed that these mixing enhancement schemes could lead to greater mixing, but often times at the cost of higher stagnation pressure losses.

There are two goals of the current study. First is to investigate the best placement of multiple injectors to obtain optimum mixing with the freestream. Although this problem could be studied computationally, the flowfield from multiple injectors presents a significant problem due to the highly three dimensional nature of the flow and the need to have a grid which can resolve the significant gradients of the complex interaction between the shocks and multiple jets. Because of these computational limitations, and the availability of the supersonic windtunnel at Rutgers the current study focuses on investigating this problem experimentally. This leads to the second goal of the study which is to create an experimental optimization methodology incorporating image processing to evaluate the mixedness, rapid prototyping to manufacture the models, and optimization techniques to suggest new geometries to evaluate. Not only does this reduce the number of experiments needed to find the best geometry for this study, but will create the infrastructure for future studies of other flowfields. Two relatively simple geometries were investigated in this study: the interaction of two circular jets with a Mach 3 free stream and the interaction of three circular jets with a Mach 3 free stream. The design variables were the relative positions of the circular injectors for the jets with the goal of optimizing the mixing of the jet. Following is a description of the optimization routine used in the study, a description of the experiments conducted, and discussion of the results.

Optimization Methodology

The first step in the experimental optimization of the mixing of multiple transverse jet injection is to select the optimization method to be used to explore the design space (i.e. placement of the injectors), evaluate the figure of merit (FOM, i.e. mixing), and propose new geometries. Once

the optimization method is found, this will help guide how initial geometries are selected, the mixing definition are evaluated, new geometries selected, and models manufactured.

Within the field of fluid dynamics there has been much interest in using optimization techniques particularly in the design of aerodynamic structures. For example, shape optimizations have been performed to redesign hypersonic inlets to cancel cowl shocks and cowl-generated compression [Shukla et al., 1996], to minimize the peak inlet Mach number [Reddy and Reddy, 1995], and to minimize total pressure loss [Munipally et al., 1995]. In general these studies are computational, using routines which can be evaluated several thousands of times with modest computer resources. As mentioned previously the current use of a computational model in the current study is prohibitive and therefore the optimization method must reflect the constraints of the experimental procedure (i.e. a limited number of experiments which can be performed).

Various optimization algorithms have been used in the past. The simplest type are gradient based methods where the differences in the value of the objective parameter over its domain are used to calculate the slope and "point" to the best location. This method has the advantage of simplicity and can always lead to a better solution, but has the disadvantage of often requiring many iterations to converge on the optimum and a difficulty of knowing before hand what the step size should be. In addition, gradient based schemes can sometimes become trapped on a portion of the domain and converge on a globally non-optimum solution. Genetic algorithms attempt to mimic the process of natural selection by starting with a collection of possible solutions. At each iteration, through processes of recombination and mutation better solutions are formed until the optimal solution is achieved. Genetic algorithms have the advantage of working well on complicated domains and being able to handle many types of optimization functions (discrete, continuous, mixed) but have the disadvantage of requiring large data sets. Random probe methods have also been used for optimization. Here a function representing the objective parameter is randomly probed over its

domain in order to locate possible optimized regions. At each iteration the domain is reduced to those areas where local optimum solutions were found. The disadvantage of this method is also the need for large data sets.

At first a gradient based method was considered to optimize a function which exactly fit the data and related the objective to the hole locations. Once the initial sets of data are taken, the design parameters and objective function data could be substituted into the function and coefficients are solved. In order for the solution to exist, the number of unknown coefficients and data sets must be equal. The problem can then be expressed in matrix form as :

$$[1, x, x^2, x^3, \dots] [A] = [F] \quad (2)$$

where the first matrix represents polynomial terms of hole positions, $[A]$ is the coefficient matrix to be solved for, and $[F]$ is the objective function vector. These steps can be performed by a Matlab program which prompts the user to enter the quantity to be optimized (i.e. mixedness definition). It then optimizes the function subject to design constraints starting from an initial guess. This initial guess is varied over some possible values for positions of the injectors. The program used a Sequential Quadratic Programming method in which a subproblem obtained by linearizing the nonlinear constraints is solved at each iteration. The routine produces a series of optimized geometries and their corresponding objective function values. These new geometries could be fabricated and tested to see how well they agree with the predicted values and also provide additional data points for a more refined optimization search (i.e. an objective function with more terms). This procedure can then be repeated until the results converge to within some tolerance.

One disadvantage of this method lies in trying to match the data exactly. In doing so, the objective function is often required to make large deviations from the general pattern produced by

the data sets giving unphysical function values. In order to avoid this, a regression analysis was used instead of a perfectly fit function. This requires more data for a polynomial of the same order, but smooths for experimental inaccuracies. Two criteria were used to determine the order of the polynomial, the number of experimental data points available in the local region inspected, and that the difference between the experimental points and the function is less than 5%. In addition, a better optimization scheme was found which makes use of both random probe and gradient methods. This technique is referred to as Simulated Annealing with Iterative Improvement --SAWI (Aly et al., 1995), and is based on the simulated annealing (SA) optimization algorithm first introduced by Kirkpatrick et al. (1983). SAWI essentially initiates the SA process to locate the neighborhood of the global optimum. Prior to the convergence of SA, the algorithm switches to random search iterative improvement, a local search method, to converge to the optimum. The Transition Parameter, a measure of where SA is in the optimization process, is used to terminate SA and initiate iterative improvement. As SA is relatively inefficient in converging to a final solution once its neighborhood has been identified, the use of a local search method significantly reduces the overall computation time (Ogot and Alag, 1993). These new geometries are fabricated and tested to see how well they agree with the predicted values and also provide additional data points for a more refined optimization search (i.e. an objective function with more terms). This procedure can then be repeated until the best geometry has been found.

Experimental Facility

The experiments were performed in the variable supersonic wind tunnel located in the Gas Dynamics and Laser Diagnostic Laboratory at Rutgers University. The tunnel operates at a stagnation temperature and pressure of 293 K and 584 kPa respectively. The nozzle is adjustable and can vary the Mach number continuously from 1.2 to 3.5, for the present study the Mach number

was set to 3.0. The injector pressure was regulated to a pressure of 412 and 390 kPa for the two and three hole injector cases resulting in a jet-to-free stream momentum flux ratio of 1.5 and 1.43, respectively. The test section has cross sectional dimensions of 102 mm by 114 mm with glass windows for optical access on three sides.

For the present study, a removable circular insert containing the injector geometry was made to fit into the wind tunnel wall. The circular insert was created using a rapid prototype process termed stereo-lithography. The stereo-lithography system (SLA 190/20) is controlled using the IDEAS CAD/CAM software package. A solid modeling program was written to automatically create the circular insert with the multiple injectors with inputs of the x-y positions, diameters, eccentricity (for elliptic geometries) and injection angle of the holes of the injector. Also the program automatically incorporated a converging transition region for each jet as well as the groove for the o-ring used to seal the circular insert. It should be noted that the jets studied in this investigation were all circular and perpendicular to the flow. The total time to create a set of 6 injectors is about 12 hours due to the curing time of the uv sensitive polymer. It could be argued that this simple geometry could be more easily manufactured using standard machining techniques, but it was desired to create a more generalized infrastructure for more complicated models to be studied in the future.

Figure 2 gives a schematic of the tunnel, laser, and camera geometry. The frequency-doubled Nd:YAG laser used for this experiment was a Spectra Physics GCR-230 with a capability of delivering 600 mJ per pulse. The pulse duration of approximately 10 ns effectively freezes the image of the turbulent motion in the flow field resulting in instantaneous measurements. The laser beam was formed into a collimated sheet by a combination of cylindrical and spherical optics. First the collimated beam (9 mm in diameter) passes through a cylindrical lens which expands the light in the lateral dimension relative to the tunnel. This is followed by a spherical lens which reduces the thickness of the laser sheet providing a thin light sheet passing spanwise through the wind tunnel.

To visualize the mixed region, the jet air was bubbled through a seeding chamber filled with ethanol. Upon mixing with the cold free stream the ethanol freezes forming small particles which mark the mixed fluid. This technique called product formation has been used to study the mixing of compressible shear layers and axisymmetric jets [Clemens and Mungal, 1992]. The scattered signal is collected using a Princeton Instruments 14-bit intensified CCD camera. The images are stored on a Pentium 100 MHZ personal computer providing camera control, laser synchronization, and image processing. All the images used in the mixing analysis described below were oriented in the spanwise direction with the flow oriented out of the page having a spanwise scale of 68 mm across the image.

Experimental Procedure

The illustration in Figure 3 gives a schematic of the experimental procedure used to optimize the injector hole positions. The first step in the process (Step i) is to select random points in the design space which "fully" represent the space and have a significant variation in the FOM. Included in the initial selection of injector positions is the base line case which is a single jet with the same exit area as the individual jets combined. In the next step (Step A) the circular inserts for each hole pattern are manufactured using stereo lithography. In general 6 inserts are manufactured simultaneously. Next each geometry is run in the wind tunnel (Step B). The experiments included recording 75 instantaneous spanwise images at three downstream locations ($x/D = 5, 11, 17$) from the centroid of the hole pattern. For image processing purposes background images and an image of a dot card was taken for each case.

Before the mixing quantities are calculated from the spanwise images they are corrected for intensity and geometric distortion (Step C). To correct for intensity, a 75 instantaneous background images without the signal from the flow were taken and averaged together. This average background

image was subtracted from each image and the images were normalized by the average maximum value. This insured that images from day to day and at different streamwise locations had the same dynamic range. Since it was desired to measure these quantities at different locations downstream, the camera and laser sheet had to be frequently moved producing a geometric bias in the size and location of the images as well as bias due to the oblique viewing angle of the camera. To correct for these geometric effects, an image of a grid of dots was taken each time the camera's position changed. An image processing program was developed which located the dots on the dot card and uses this information to realign the images correcting for geometric distortions due to image size changes and viewing at an oblique angle.

From the experimentally obtained spanwise images three different figures of merit were calculated with a fourth parameter calculated by combining all three together and weighting them equally (Step D). Figure 4 illustrates two of these parameters the mixing area (A_p) and jet penetration depth (y_p). The mixing area represents the amount of the jet that has been entrained into the freestream. The mixing area for each instantaneous image was calculated by summing the number of pixel which are above 5% of the average maximum value. These values are then averaged and converted to square mm. The repeatability of the mixing area to day to day variations and tunnel conditions variation was found to be less than 2%. The jet penetration represents the ability of the jet to intrude into the free stream away from the surface. The penetration distance is found by taking the centroid of the mixing area images and calculating the distance from the centroid to the surface. Another parameter that represents the ability of the transverse jets to mix with the freestream is the growthrate. The growthrate represents how quickly the jet is entrained into the free stream as the jet develops downstream. From the images a growthrate (dA_p/dx) can be calculated from the slope of a linear curve fit through the three average mixing areas calculated at each streamwise location. It should be noted that there are many other definitions which could be used to define the "best"

mixing and it may depend on characteristics specific to the application. Also any one of these three definitions may be best suited, therefore the hole pattern was optimized for each individual mixing definition (i.e. mixing area, penetration, and growthrate). It could be argued, however, that all these mixing definitions are important and so a combined figure of merit is calculated representing a forth mixing definition. The combined figure of merit (CFOM) is given by

$$CFOM = \sqrt{\frac{A_p^2}{A_{p0}^2} + \frac{y_p^2}{y_{p0}^2} + \frac{[A_p/dx]^2}{[A_{p0}/dx]^2}} \quad (3)$$

where each mixing definition is normalized by its value for the baseline single circular jet case denoted by the subscript 0. It should be noted that individual mixing definitions were weighted equally in the present study, but it may be desirable in the future to consider one or more definitions more important. Also the CFOM is normalized to unity for the best case.

After the data is taken from each manufactured jet geometry and the four mixing definitions are evaluated the SAWI optimization algorithm is initiated (Step E). Within the SAWI algorithm described above the large design space is reduced around promising regions and a least square polynomial is fit to the data in this region. Note again that the polynomial was made to be the lowest order possible to describe the “optimal region” (instead of the entire data set) as defined by the residue between the function and experimental data (in general the residue was less than 5% in these regions). Candidate designs (i.e. jet orientations) for each FOM to be manufactured were selected (Step F) if they were significantly different (greater than 2%) than designs already evaluated as well as higher than the baseline case. Therefore new designs were evaluated which may be less than the current optimum, but enable the design space to be checked in promising regions. The new designs were then manufactured (Step A) and the process was repeated. Termination of the optimization (Step G) was decided when no significantly different points were suggested by the SAWI algorithm

and when the predicted values and actual values of the FOM were within 5%. Following is the results of using this experimental methodology to optimize two and three injector geometries for mixing.

Two Injector Optimization

In order to test the experimental methodology and optimization algorithm the first case studied was a two injector geometry. This allowed the design space to be visualized since the number of design parameters with constraints is reduced to two with the FOM evaluated (i.e. mixing area, penetration, or the combined figure of merit) as the third dimension of the surface. Also, it is much easier to understand the physics of the mixing since there is less interaction than the three jet geometry. For the two injector “proof of concept” experiments the growthrate was eliminated from the analysis so that the number of tunnel runs could be reduced with only one streamwise location needed to calculate the mixing area and penetration depth. Images were taken of the single injector baseline case so that the mixing definitions (i.e. figure of merit, FOM) could be calculated. The diameter of the single injector case had the same total exit area as the two jet geometries resulting in a jet diameter of 3.46 mm for the single jet and 2.54 mm for the two jets.

Four variables are needed to describe the locations of the two holes on a Cartesian plane (i.e. x_1, z_1, x_2, z_2) as shown in Figure 5. The problem is constrained in that the holes must be located on the injector disk with a radius (R_{disk}) of 12.2 mm. The optimization problem then can be represented as follows:

$$\begin{aligned} \text{maximize } f(x) = & a_0 + a_1 x_1 + a_2 z_1 + a_3 x_2 + a_4 z_2 \\ & a_7 x_1^2 + a_8 z_1^2 + a_9 x_2^2 + a_{10} z_2^2 + \dots \end{aligned}$$

Subject to:

$$x_1^2 + z_1^2 \leq R_{disk}^2$$

$$x_2^2 + z_2^2 \leq R_{disk}^2$$

$$x_1 + x_2 = 0$$

$$z_1 + z_2 = 0$$

where $f(x)$ is the figure of merit to be optimized and a_i the coefficients of the polynomial. The x and z centroid of the hole pattern was constrained to be at the center of the disk. The z centroid was constrained to the center to prevent physical redundancy and the x centroid to allow for comparison between cases. In addition, since interchanging x with $-x$ or z with $-z$ produces the same physical geometry, only even terms are needed to fit the data. This allows for odd terms to be removed from the function and higher order curve fits can be used. These equalities reduce the problem to that of two design variables.

Figure 6 illustrates the progression of the mixing definitions of mixing area, penetration, and combined figure of merit (CFOM) for each iteration step including the single jet baseline case at iteration number zero. Initially, eight geometries were chosen which would span the domain appropriately with the SAWI optimization routine suggesting at least one new geometry for each FOM at successive iterations. The value of the FOM for all geometries are shown at each iteration even though some are used to optimize another FOM. In all cases the best figure of merit was found at the last iteration although at the intermediate iteration (iteration #2) the figure of merit dipped down slightly. This is due to the SAWI algorithm and the fact that most optimum values are only evaluated if they are significantly different than geometries already evaluated. Table 1 gives the baseline (single injector) and optimized figure of merits for mixing area, penetration, and CFOM.

In all mixing definitions the FOM was improved (82% for mixing area and 43% for the CFOM), but penetration had the least improvement of only 7.3%.

Figures 7 and 8 give the physical injector layout, average image, and one instantaneous image for the single jet baseline case, best mixing area geometry, penetration geometry, and CFOM geometry. The single jet case in Figures 7a and 7b show typical transverse jet injection characteristics with counter-rotating vortex pair clearly visible in the average and instantaneous images. The injector geometry which optimizes the best mixing area (Figure 7c and d) arranges the injectors in a line perpendicular to the flow. This is physically reasonable since holes spaced far apart would behave more like two independent jets producing greater total mixing area than when the jets interfere greatly. It should be noted however that the spacing was not to the maximum edge of the design domain suggesting there is still some interaction. The geometry and resulting images giving the optimized penetration depth is shown in Figure 8a,b. This geometry has the injectors lined up with the flow direction and wide apart, but again not to the maximum edges of the design domain which seems reasonable since it would be expected that the wake and counter-rotating for the first jet would enhance the penetration depth of the second jet. The average and instantaneous images show a similar flow pattern to the single jet case with the counter rotating vortices clearly observed. Figure 8c,d show the geometry for the CFOM and resulting average and instantaneous image. This geometry is the same as the mixing area geometry for reasons which will be discussed shortly.

Since there are only two design variables for the two injector geometry, a surface plot can be made for each mixing definition with the FOM used for the third dimension. Figure 9 gives the surface plot for mixing area, penetration, and CFOM. A sixth order polynomial was found to fit the data sets with less than 5% standard deviation from the function and actual data. Upon performing the optimization on the individual area function, the results gave optimized geometries primarily

with the holes in a line perpendicular to the flow. This can be seen graphically in Figure 9a where the objective function grows as the outside edges of the disk are reached ($z=\pm 12.2$ mm). Similarly, when penetration depth was the individual objective function, the results indicated that the best geometry was for injectors lined up along the flow direction as can be seen in Figure 9b where two distinct peaks occur for $z=0$ and $x=\pm 12.2$ mm. As mentioned previously these geometries allow the jets to interfere constructively and reinforce each other to penetrate the free stream. As seen in Figure 9c the CFOM optimization gave similar results and function shape to the area optimization. This is due to the fact that the area changes by up to 82% while the penetration changes by at most 7.3%. Since it was assumed that these two objectives were equally important in defining mixing, the figure of merit calculation weighs these two objectives equally. Therefore, the changes in area outweigh the changes in penetration and dictate the CFOM.

Three Injector Geometry

Now that the optimization algorithm has been shown to give reasonable results for the two injector geometry, the complexity of the geometry and analysis was increased. The second optimization experiment involved finding the location of three injectors. Also the growth rate was incorporated as an additional FOM evaluated individually and added into the CFOM. This geometry proved to be more interesting since the effects of jet interactions were enhanced. Here, the geometries were no longer symmetric with respect to the z -axis (ie. x can not be interchanged with $-x$) so odd terms in x were required. Using similar constraints as in the two-hole case and forcing one of the holes to lie along the x -axis ($z=0$), the problem could be reduced to three variables. This still presents the problem of plotting and viewing the data since three variables are needed to represent the geometries and a fourth to represent the objective function value.

Six variables were needed to describe the locations of the three holes on a Cartesian plane

(i.e. $x_1, z_1, x_2, z_2, x_3, z_3$) as shown in Figure 10. The diameter of the single injector case had the same total exit area as the three jet geometries resulting in a jet diameter of 6.35mm for the single jet and 3.67mm for the three jets. The problem is constrained in that the holes must be located on the injector disk with a radius (R_{disk}) of 10.6 mm and that the x and z values must be in ascending order. This last constraint prevents redundancy in the configurations since interchanging holes would produce the same result. The optimization problem then can be represented as follows:

$$\begin{aligned} \text{maximize } f(x) = & a_0 + a_1 x_1 + a_2 z_2 + a_3 x_2 + a_4 z_2 + a_5 x_3 + a_6 z_3 + \\ & a_7 x_1^2 + a_8 z_1^2 + a_9 x_2^2 + a_{10} z_2^2 + a_{11} x_3^2 + a_{12} z_3^2 + \dots \end{aligned}$$

Subject to:

$$x_1 \leq x_2 \leq x_3 \quad \text{OR} \quad z_1 \leq z_2 \leq z_3$$

$$x_1^2 + z_1^2 \leq R_{disk}^2$$

$$x_2^2 + z_2^2 \leq R_{disk}^2$$

$$x_3^2 + z_3^2 \leq R_{disk}^2$$

$$x_1 + x_2 + x_3 = 0$$

$$y_1 + z_2 + z_3 = 0$$

$$z_2 = 0$$

To simplify this procedure, one hole is constrained to be on the centerline ($z_2=0$), making one of the z values zero. Also, the x and z centroid of the holes were constrained to be at the center of the disk. The z centroid was constrained to the center to prevent physical redundancy and the x centroid to allow for comparison between cases. These three equalities reduce the problem to that of three design parameters. Also, an addition constraint was chosen based on the difference of new geometries from those already evaluated. If the difference of the new geometries was less than 2% other geometries were chosen. This means that the last iteration may not show improved values, but the diversity of the geometries evaluated is always increased.

Table 2 gives the baseline (single injector) and optimized figure of merits for mixing area,

penetration depth, growthrate, and the CFOM. In all mixing definitions the figure of merit was improved (66% for mixing area, 12% for growthrate, and 9% for the CFOM), but again penetration had the least improvement of only 2%. Figures 11 and 12 give the physical injector layout, average image, and one instantaneous image for the single jet baseline case, best mixing area geometry, penetration geometry, and CFOM geometry. The single jet case in Figure 11a and 11b again show a typical transverse jet injection characteristics as described previously. The injector geometry which optimizes the best mixing area (Figures 11 c,d) arranges the injectors in a line spread out, but not totally perpendicular to the flow like the two hole geometry. This indicates that the interaction between the jets is much more important for the three injector geometry. The geometry and resulting images giving the optimized penetration depth is shown in Figures 12a,b. This geometry, similar to the two hole geometry, has the injectors lined up with the flow direction with a wide separation. Again this geometry seems reasonable since the wake and counter-rotating for the first jet would enhance the penetration depth of the second jet. The average and instantaneous images how the existence of counter rotating vortices clearly observed. Figures 12c,d show the geometry for the best growth rate and resulting average and instantaneous image. The three injectors in this geometry is aligned with the flow to maximize the interaction between the jets. It is also interesting to note that the average and instantaneous images clearly show the counter-rotating vortex structure which may be essential to maximize the growthrate. The growth rates over the domain had the most variation of the FOM's evaluated ranging from 0 to 1.25 mm²/mm. Physically, this would indicate that injectors separated by some finite distance with pronounced interference effects provide the optimum growthrate. The CFOM has the same geometry as the growthrate due to the fact that the growthrate has much higher variation between geometries and therefore has the greatest effect since the FOM are equally weighted.

The main difficulty involved with the three hole case was getting a physical picture of the

objective function due to the presence of the third independent variable, thus becoming a four dimensional problem. As a result interpretation for the design domain is much more difficult.

Conclusions

Experiments were performed on transverse injection into a supersonic flow using optimization techniques. The first experiment involved finding the best position for two equal diameter injectors which gave the best cross-sectional area and penetration depth, both individually and combined. It was found that changes in area outweighed changes in penetration depth and was the greatest when the two injectors were placed far apart in a line perpendicular to the flow. The second experiment involved finding the best position for three equal diameter holes with the added objective of optimum growth rate. Although results could not be as easily interpreted as in the two injector case, they suggest that optimum solutions combining these three parameters exist for geometries with injectors at intermediate distances from the center and give improved mixing as compared to an equivalent single hole. Further research beyond the scope of this preliminary investigation will involve studying other optimization routines applied to experimental analysis, consider other mixing definitions experiments which would explain the improvements in mixing, and more complicated geometries better utilizing the benefits of stereo-lithography. Aspects of this research program has been accepted for presentation in ASME and AIAA fluids conferences.

Acknowledgments

The principle investigator would like to thank graduate students Bradford Price and Robert Murray for their help in completing this research as well as the support of researchers Dr. Abdollah Nejad, Dr. Mark Gruber, and Dr. Diana Glawe at Wright Laboratory for their input and assistance in outlining this research program and input interpreting the results.

References

- Aly, S., M.M. Ogot and R. Pelz, "An Improved Simulated Annealing Algorithm" Proceedings of the 21st ASME Design Automation Conference, Sept. 1995, Boston, MA.
- Clemens, N.T., and Mungal, M.G., "Effects of Sidewall Disturbances on the Supersonic Mixing Layer," *AIAA Journal*, Vol. 8, No. 1, pp. 249 - 251, 1992.
- Cox, S.K., Fuller, R.P., Schetz, J.A., and Walters, R.W., "Vortical Interactions Generated by an Injector Array to Enhance Mixing in Supersonic Flow," AIAA 94-0708, 1994.
- Donohue, J.M., McDaniel, J.C., and Haj-Hariri, H., "Experimental and Numerical Study of Swept Ramp Injection into a Supersonic Flowfield," *AIAA Journal*, Vol. 32, No. 9, pp. 1860-1867, 1994.
- Fuller, R.P., Wu, P.-K., Nejad, A.S., and Schetz, J.A., "Fuel-Vortex Interactions of Enhanced Mixing in Supersonic Flow," AIAA 96-2661, 1996.
- Gruber, M.R., Nejad, A.S., Chen, T.H., and Dutton, J.C., "Mixing and Penetration of Sonic Jets in a Mach 2 Free stream," *Journal of Propulsion and Power*, Vol. 11, No. 2, pp. 315-323, 1995.
- Gruber, M.R., Nejad, A.S., Chen, T.H., and Dutton, J.C., "Large Structure Convection Velocity Measurements in Compressible Transverse Injection Flowfields," *Experiments in Fluids*, Vol. 22, pp. 397-407, 1997.
- Hartfield, R.J., Hollo, S.D., and McDaniel, J.C., "Experimental Investigation of a Supersonic Swept Ramp Injector Using Laser-Induced Iodine Fluorescence," *Journal of Propulsion and Power*, Vol. 10, No. 1, 1994.
- Hermanson, J.C. and Winter, M., "Mie Scattering Imaging of a Transverse, Sonic Jet in a Supersonic Crossflow," *AIAA Journal*, Vol. 31, No. 1, pp. 129-132, 1993.
- Kirkpatrick, S. And C.D. Gelatt, "Optimization by Simulated Annealing," *Science*, 220(4598), pp. 671-680, 1983.
- Lee, M.P., McMillin, B.K., Palmer, J.L., and Hanson, R.K., "Planar Fluorescence Imaging of a Transverse Jet in a Supersonic Crossflow," *Journal of Propulsion and Power*, Vol. 8, No. 4, pp. 729-735, 1992.
- McCann G.J., and Bowersox, D.W., "Experimental Investigation of Supersonic Gaseous Injection into a Supersonic Free stream," *AIAA Journal*, Vol. 34, No. 2, 1996.
- Munipalli, R., Wadawadigi, G., Anderson, D.A., and Wilson, D.R., "Application of Optimization Techniques to Inlet Design," AIAA 95-1824, 1995.
- Ogot, M.M. and S. Alag "An Effective Mixed Annealing/Heuristic Algorithm for Kinematic Mechanical Design," *ASME Journal of Mechanical Design*, Vol. 117, No. 3, pp. 409-418, 1995.
- Reddy, E.S., and Reddy, D.R., "Aerodynamic Shape Optimization of a subsonic Inlet Using 3-D Euler Computation," AIAA 95-2757, 1995.
- Schetz, J.A., and Billig, F.S., "Penetration of Gaseous Jets Injected into a Supersonic Stream," *Journal of Spacecraft and Rockets*, Vol. 3, No. 11, pp. 1658-1665, 1966.
- Shukla, V., Gelsey, A., Schwabacher, M., Smith, D., and Knight, D., "Automated Redesign of the NASA P8 Hypersonic Inlet Using Numerical Optimization," AIAA 96-2549, 1996.

Table 1. Comparison of the optimized two injector geometries with the baseline case

Model	(-)z1 [mm]	(-)x1 [mm]	(-)z2 [mm]	(-)x2 [mm]	mixing area [mm ²]	penetration [mm]	CFOM	% increase of FOM
Single injector	0	0	0	0	322	9.82	0.698	*
Optimized for mixing area	11.76	0	-11.76	0	588	8.62	1	82%
Optimized for penetration	*0	9.31	0	-9.31	457	10.6	0.88	7.3%
Optimized for CFOM	11.76	0	-11.76	0	588	8.62	1	43%

Table 2. Comparison of the optimized three injector geometries with the baseline case

Model	(-)z1 [mm]	(-)x1 [mm]	(-)z1 [mm]	(-)x2 [mm]	(-)z3 [mm]	(-)x3 [mm]	Mixing area [mm ²]	penetration [mm]	growth rate [mm ² /mm]	CFOM	% incr. of FOM
Single injector	0	0	0	0	0	0	469	15.3	0.0443	0.949	*
Optimized for mixing area	-9.61	-4.73	0	-0.58	9.61	5.32	778	9.48	0	0.982	66%
Optimized for penetration	0	-11	0	0	0	11	566	15.6	0.0161	0.887	2%
Optimized for growthrate	0	-4.89	0	10.9	0	-6.11	511	14.6	0.0495	1	12%
Optimized for CFOM	0	-4.89	0	10.9	0	-6.11	511	14.6	0.0495	1	9%

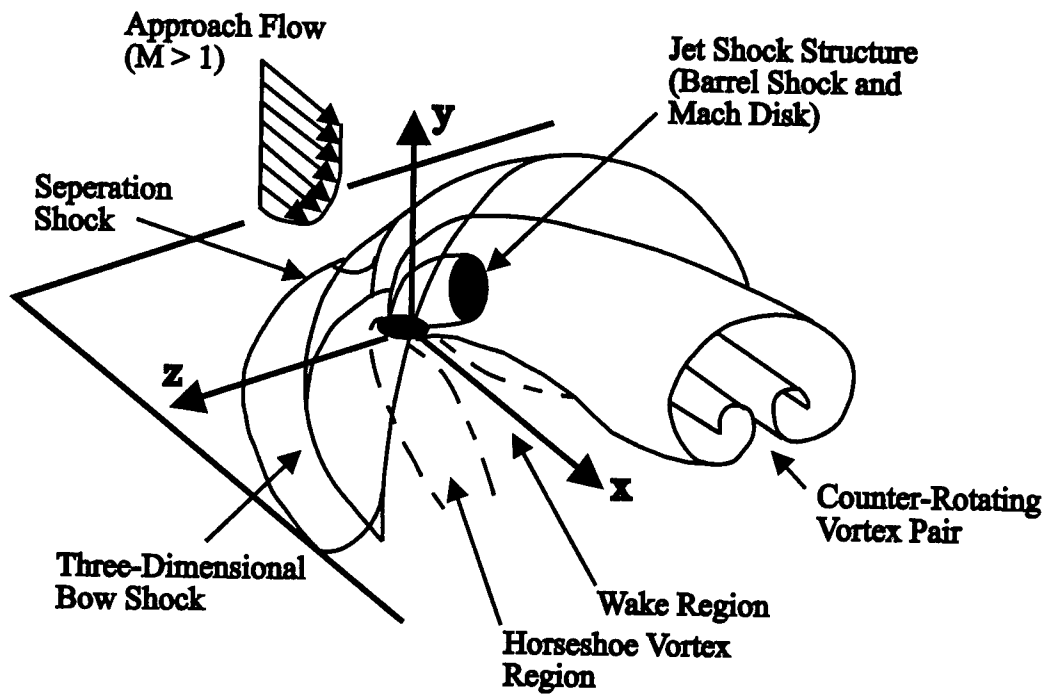


Figure 1. Perspective view of the mean transverse injection flowfield.

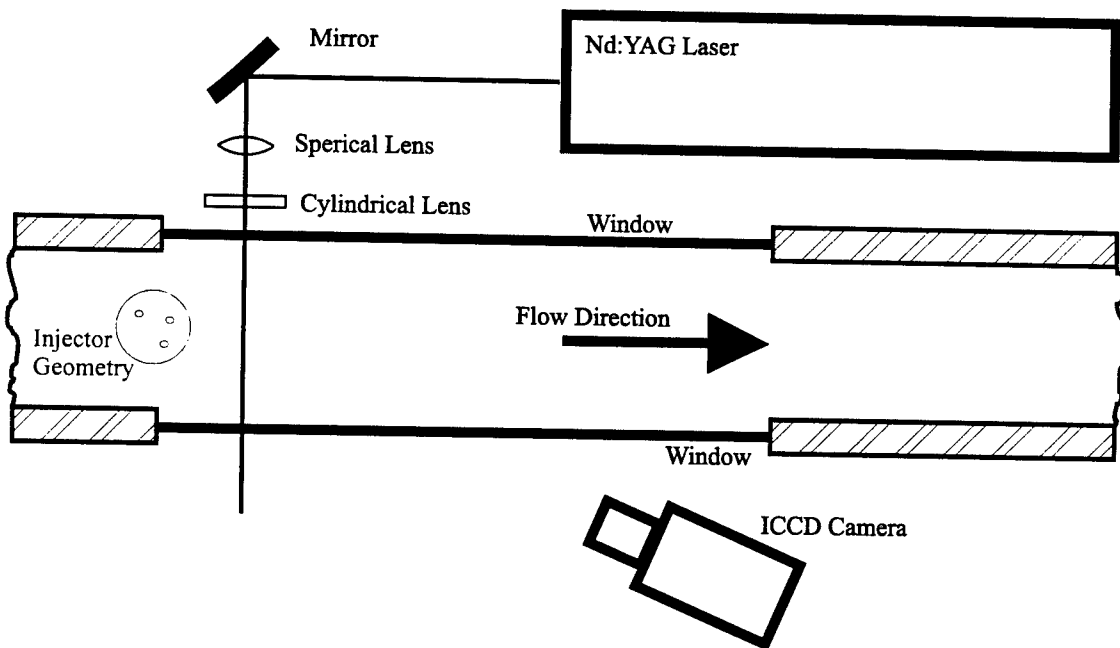


Figure 2. Plan view of the tunnel and optical arrangement for measurements of a jet in a supersonic cross flow.

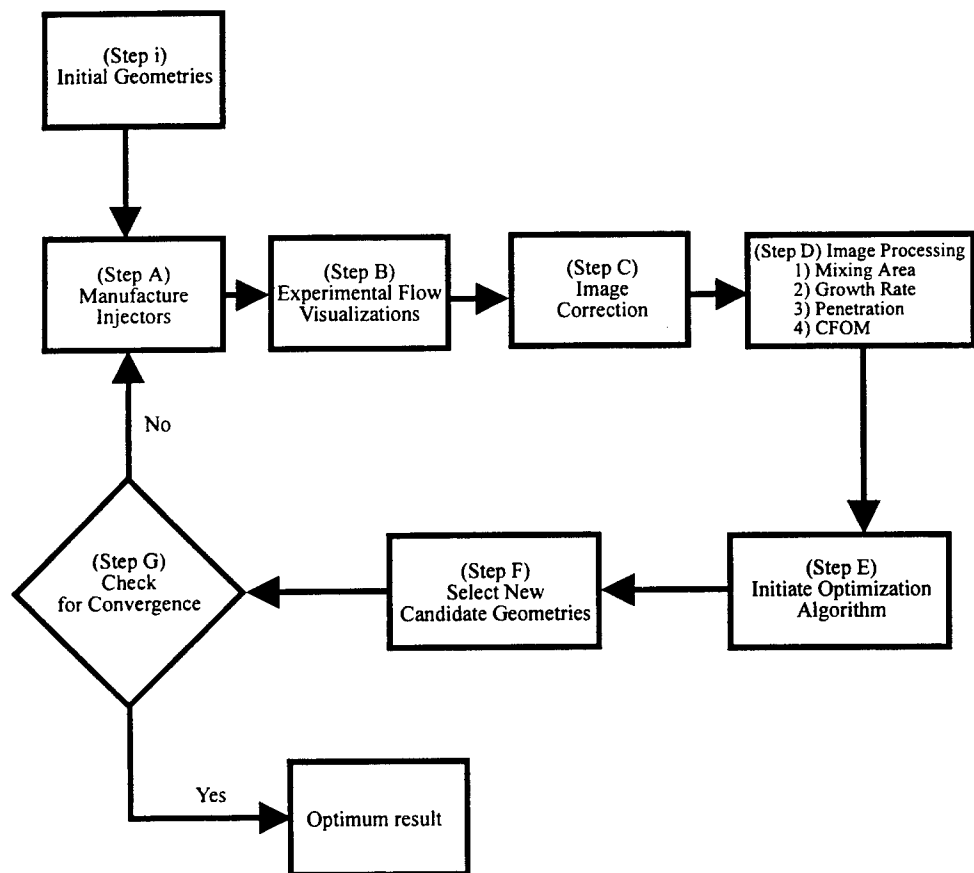


Figure 3. Schematic of optimization procedure.

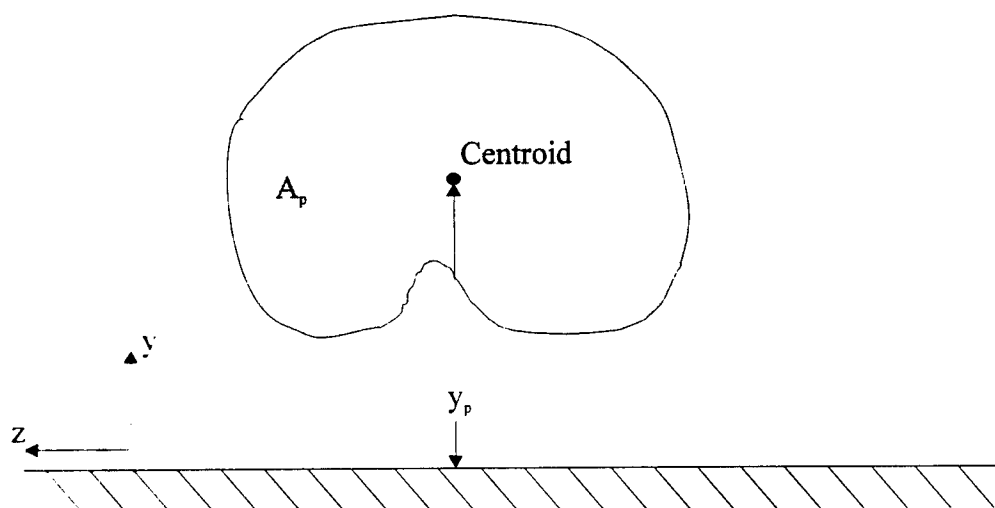


Figure 4. Schematic of mixing area (A_p) and penetration (y_p) definitions.

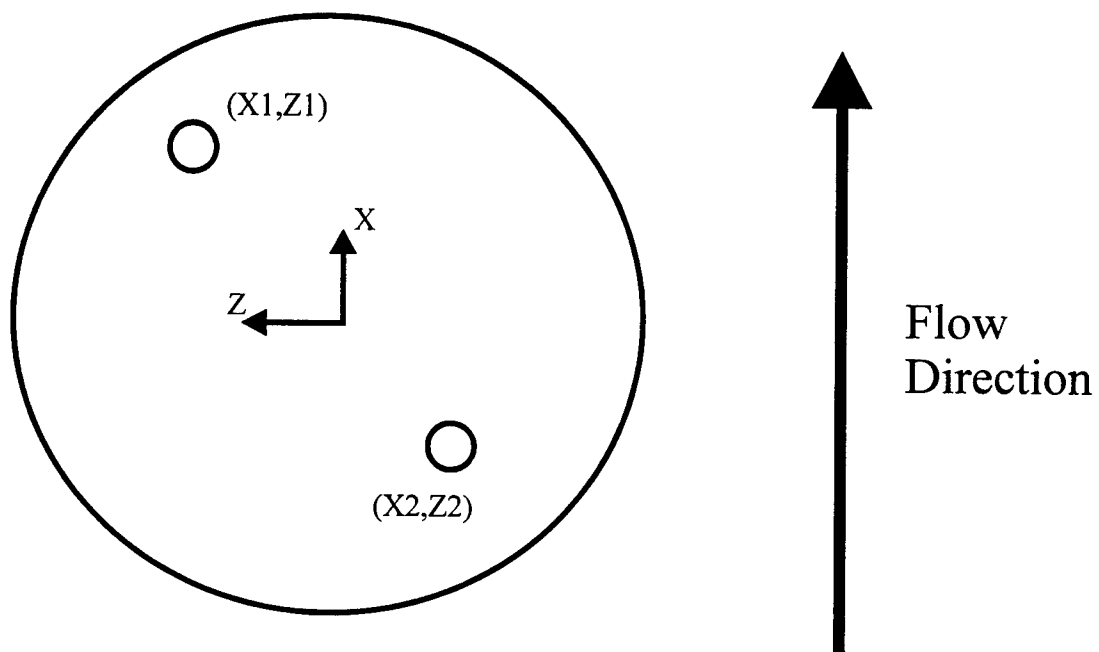


Figure 5. Two injector geometry.

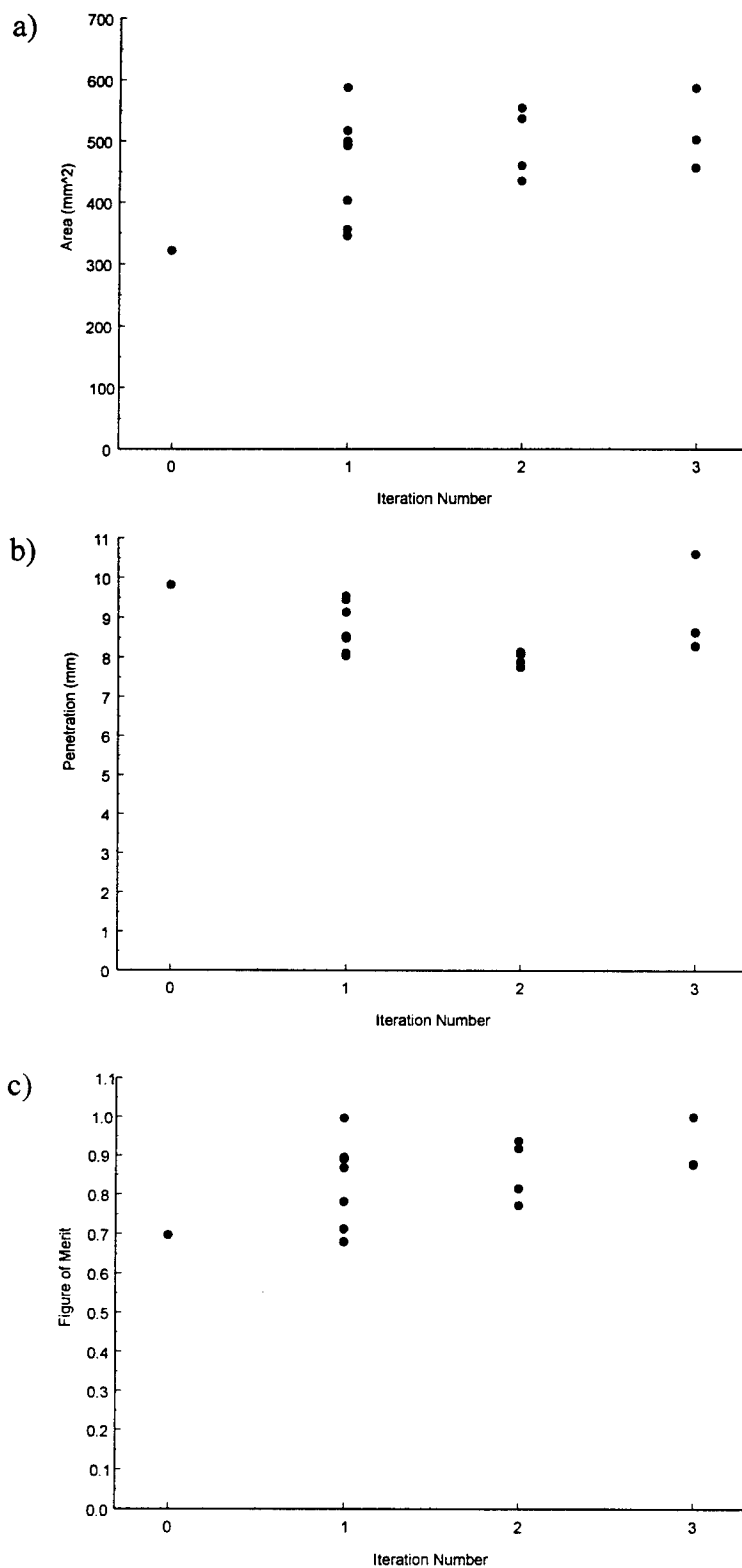


Figure 6. Evolution of average mixing area (a), penetration depth (b), and combined figure of merit (c) for two injector geometry.

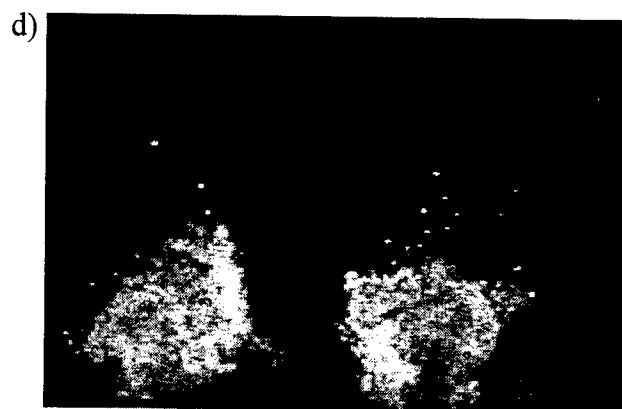
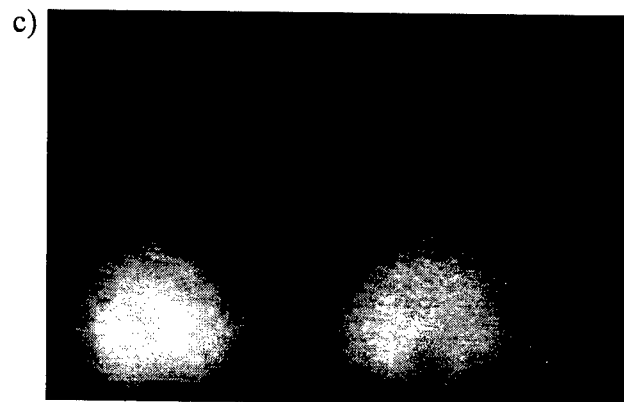
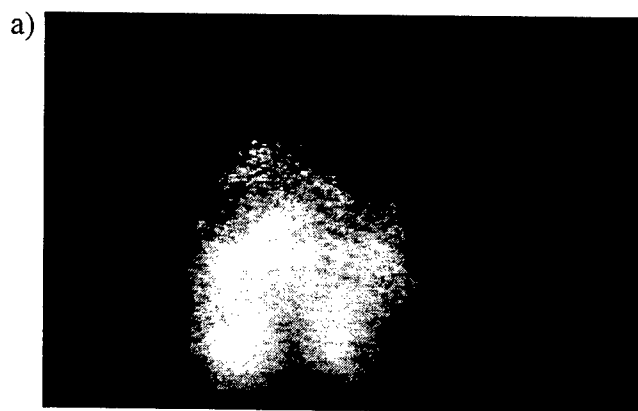
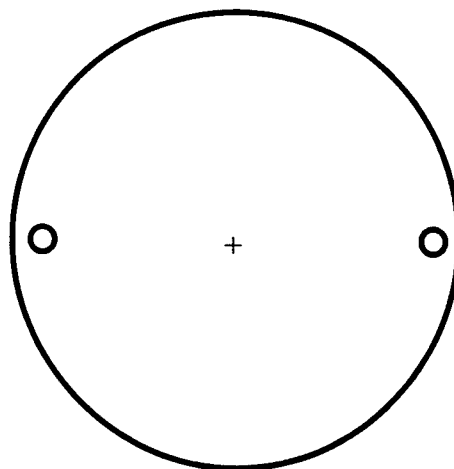
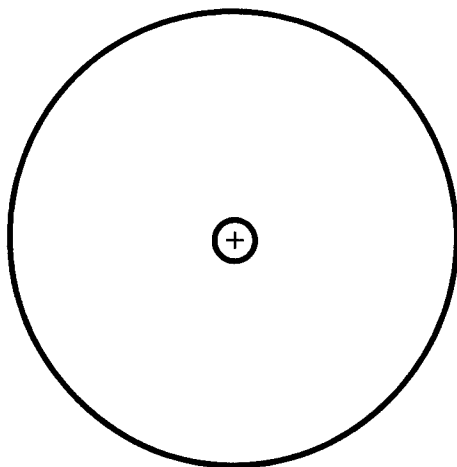


Figure 7. Spanwise view images for the averaged (a) and instantaneous (b) single injector geometry and averaged (c) and instantaneous (d) two injector geometry for best mixing area.

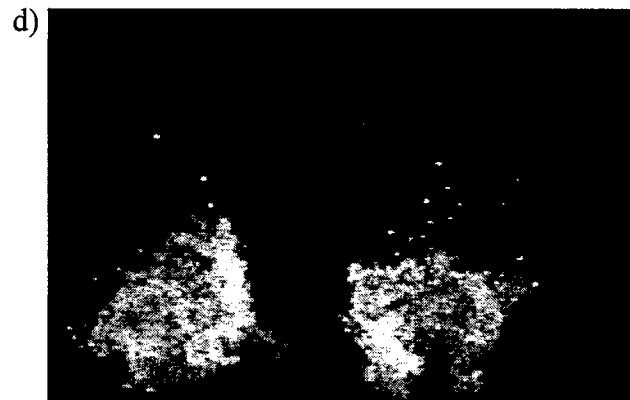
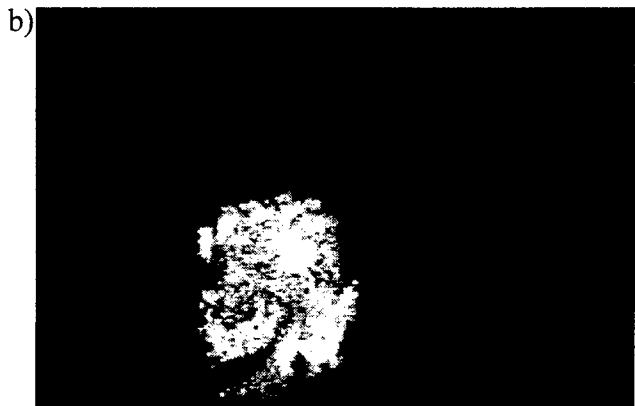
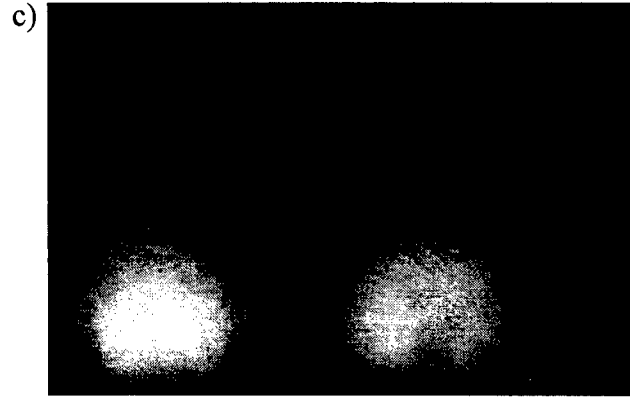
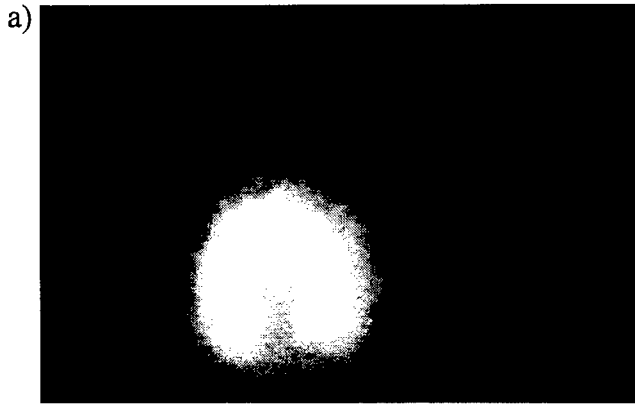
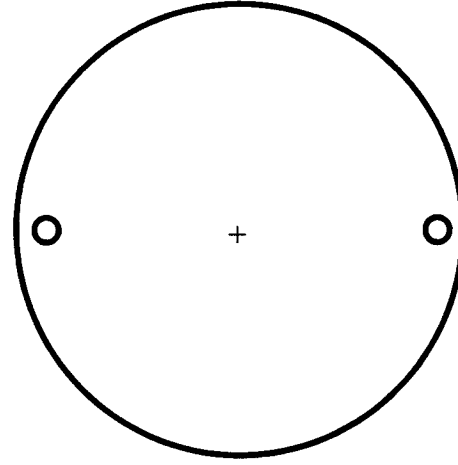
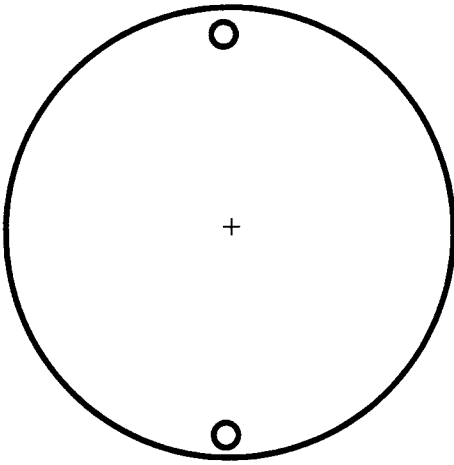


Figure 8. Spanwise view of average and instantaneous images of the two injector geometry for best penetration depth (a,b) and CFOM (c,d).

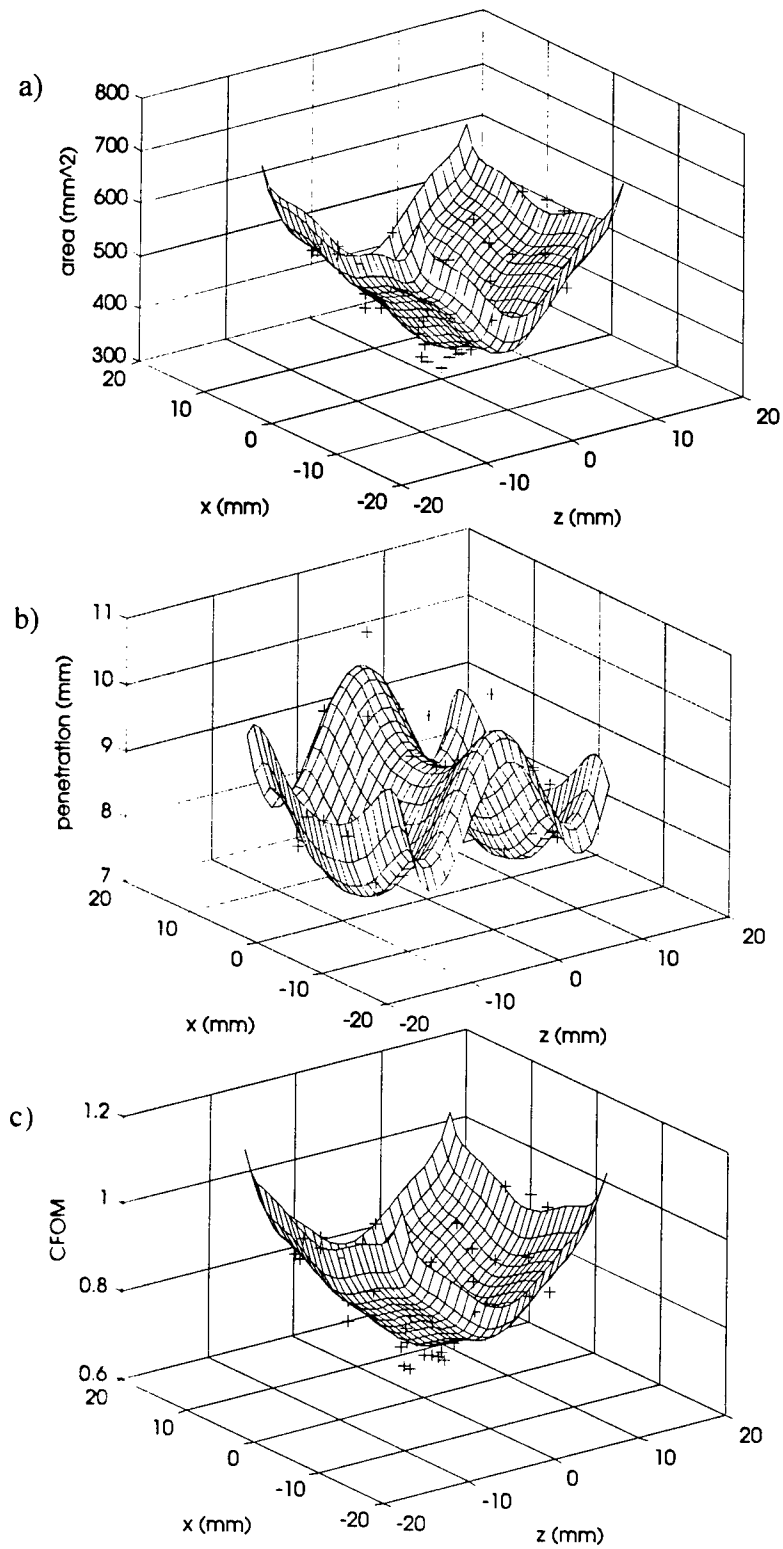


Figure 9. Contour plots of average mixing area (a), penetration depth (b), and CFOM (c) for the two injector geometry.

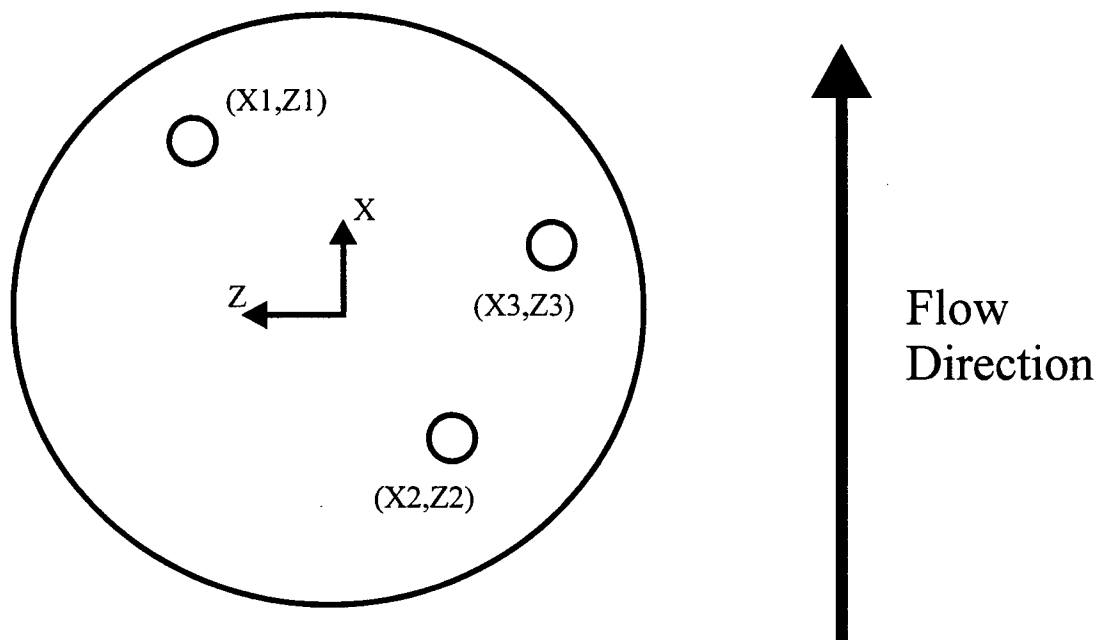


Figure 10. Three injector geometry.

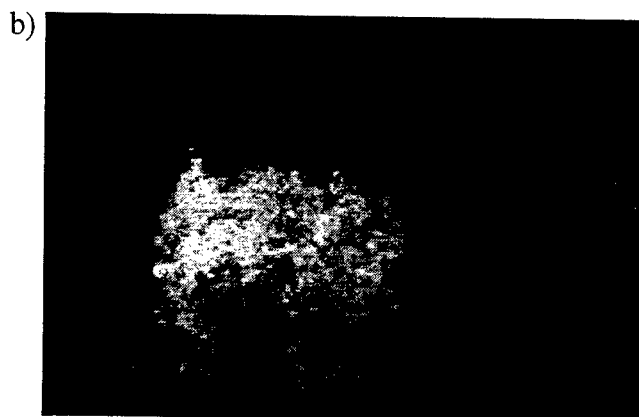
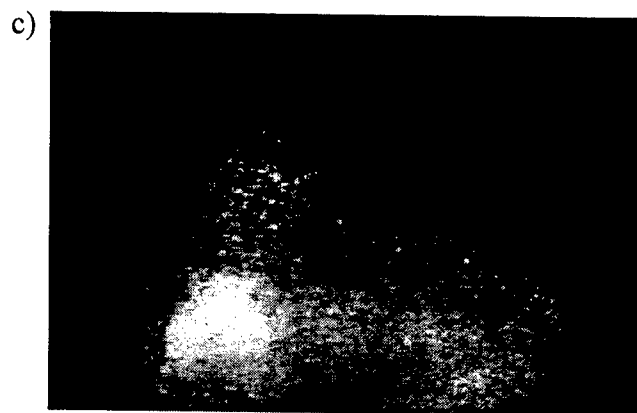
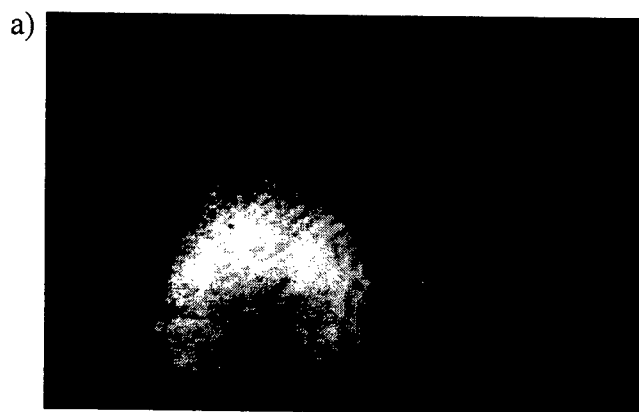
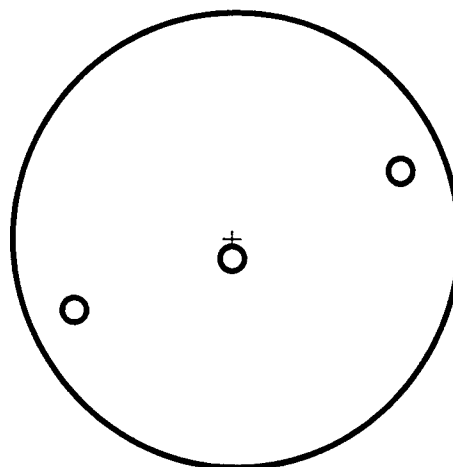
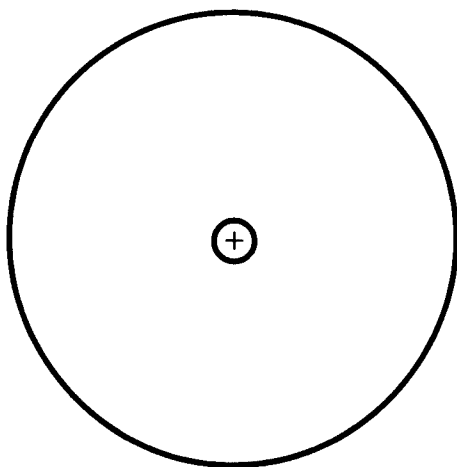


Figure 11. Spanwise view images for the averaged (a) and instantaneous (b) single injector geometry and averaged (c) and instantaneous (d) three injector geometry for best mixing area.

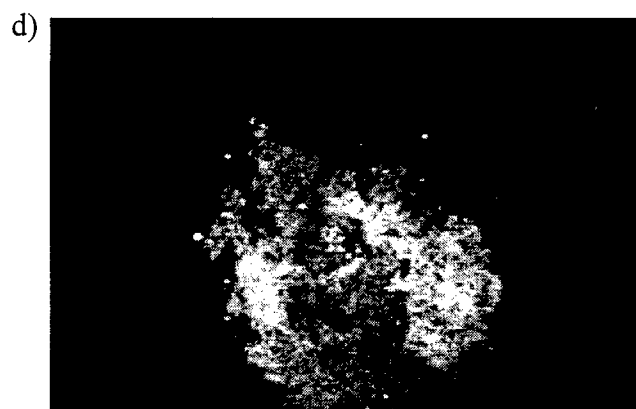
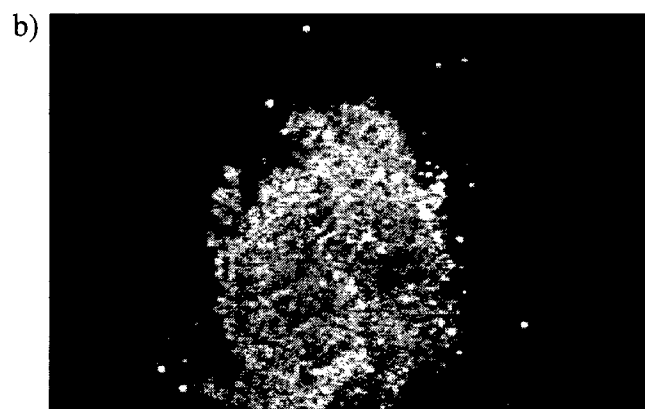
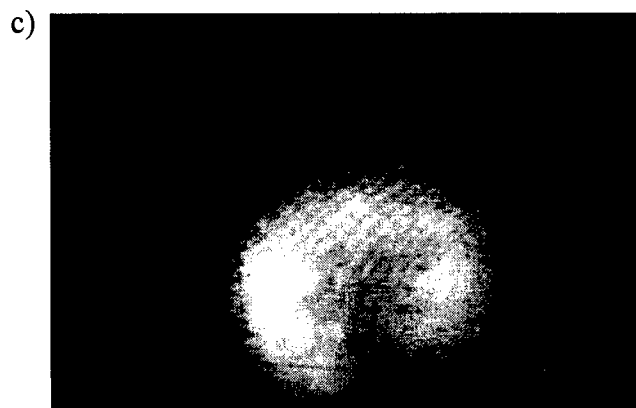
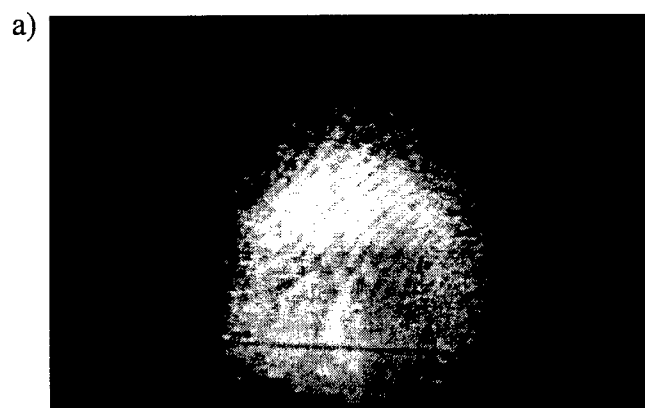
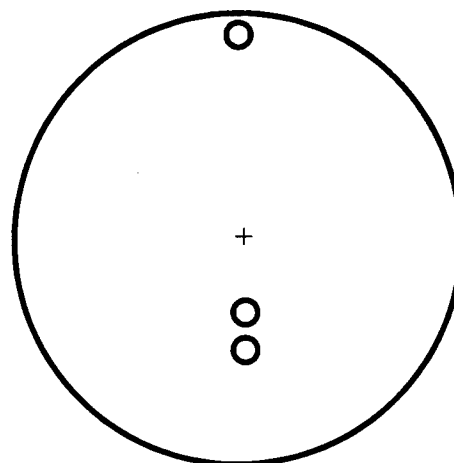
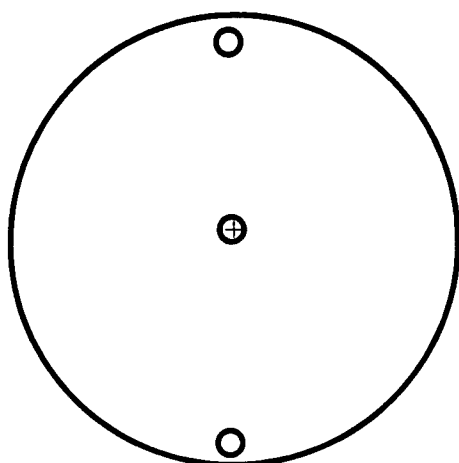


Figure 12. Spanwise view of average and instantaneous images of the three injector geometry for best penetration depth (a,b) and both growth rate and CFOM (c,d).

Appendix A

Following is the complete data of all the experiments performed for two injector and three injector geometries. The injector positions are given in coordinates which are shown in Figures 5 and 10.

Table 3. Two Injector Optimization

(-)z1 [mm]	(-)x1 [mm]	(-)z2 [mm]	(-)x2 [mm]	mixing area [mm ²]	penetration [mm]	CFOM
0	0	0	0	321.9348	9.8209	0.6975
0.50715	1.89385	-0.50715	-1.89385	356.1283	9.1281	0.7126
1.89385	0.50715	-1.89385	-0.50715	345.8058	8.4765	0.6796
1.77625	6.62725	-1.77625	-6.62725	403.8702	9.525	0.782
3.11885	3.11885	-3.11885	-3.11885	492.9022	8.529	0.8681
6.62725	1.77625	-6.62725	-1.77625	517.4183	8.0311	0.8893
6.58315	6.58315	-6.58315	-6.58315	500.6442	8.0996	0.8681
3.0429	11.3582	-3.0429	-11.3582	495.4829	9.4375	0.8948
11.3582	3.0429	-11.3582	-3.0429	587.7408	8.5087	0.9966
5.4047	4.2238	-5.4047	-4.2238	460.9462	8.1305	0.8157
10.2851	5.70115	-10.2851	-5.70115	554.8518	7.8801	0.9376
5.6203	3.9347	-5.6203	-3.9347	436.3488	7.7414	0.7733
7.62685	5.341	-7.62685	-5.341	537.7105	8.0662	0.9179
11.76	0	-11.76	0	588.3859	8.6243	1
0	9.31	0	-9.31	457.4184	10.5946	0.8798
6.5758	9.74855	-6.5758	-9.74855	503.87	8.2798	0.8767

Table 4. Two Injector Optimization

(-)z1 [mm]	(-)x1 [mm]	(-)z1 [mm]	(-)x2 [mm]	(-)z3 [mm]	(-)x3 [mm]	Mixing area [mm ²]	penetration [mm]	growth rate [mm ² /mm]	CFOM
0	0	0	0	0	0	469.982	15.3197	0.0443	0.9488
-9.525	5.4991	0	-10.998	9.525	5.4991	593.009	11.5623	0.0152	0.8271
-4.7625	-2.7495	0	5.4991	4.7625	-2.7495	495.401	14.4107	0.0222	0.821
-4.7625	2.7495	0	-5.4991	4.7625	2.7495	494.095	12.4507	0.0265	0.7982
-10.998	0	0	0	10.9982	0	548.167	11.747	0.0162	0.7903
0	-10.998	0	0	0	10.9982	566.495	15.5937	0.0161	0.8868
-3.9992	0	0	0	3.9992	0	493.573	14.0191	0.0168	0.7906
0	-3.9992	0	0	0	3.9992	471.342	13.5852	0.0264	0.8027
0	-5.4991	0	-5.4991	0	10.9982	655.215	13.8217	0.0181	0.9368
-9.6088	5.3492	0	-1.8974	9.6088	-3.4519	644.171	9.9507	0.0105	0.8409
0	-4.887	0	10.9982	0	-6.1112	510.895	14.5661	0.0495	1
-8.824	-3.4595	0	-3.1064	8.824	6.5659	672.593	10.5518	0.0343	0.9675
-4.9708	-9.812	0	0	4.9708	9.812	584.406	13.6923	0.0383	0.9634
-2.0091	-8.0086	0	-2.8042	2.0091	10.8128	570.999	14.04	0.0131	0.8492
-2.032	-2.8321	0	-2.8321	2.032	5.6642	578.115	11.6169	0.0169	0.8186
-4.638	-0.414	0	-5.8572	4.638	6.2713	667.865	9.8919	0.0057	0.8579
-6.162	-9.05	0	9.8908	6.162	-0.8407	684.751	11.6782	0	0.9049
-9.6088	-4.7371	0	-0.5842	9.6088	5.3213	778.068	9.4794	0	0.9818
-0.5715	-10.49	0	7.6048	0.5715	-2.8854	614.181	13.2856	0	0.8662

Appendix B

Although not part of the optimization procedure another geometry that was tested was the effect of crossing oblique shock waves on the mixing of a transverse jet injected in a supersonic free stream. This geometry is illustrated in Figure B1. The crossing shocks were believed to be possible mixing enhancement mechanisms due to the work conducted with side wall shock enhancement in planar mixing layers and the existence of counter-rotating vortices in the boundary layer in these flowfields. The flow conditions, seeding, mixing definitions, and optical arrangement were the same as the single injector geometry described previously for the three injector optimization. The start of the shocks was approximately 132 mm before the center of the circular jet and the two wedges were inclined at 7° relative to the free stream. Thus the crossing shocks were formed before the injector location.

Table B1 gives the mixing area, penetration depth, and CFOM of merit for the single jet with and without the crossing shocks. As seen here the crossing shocks tend to suppress the mixing for all definitions instead of increasing them. The reason for this is evident when observing the average and instantaneous images given in Figure B2. The counter rotating vortex tends to keep the jet from injecting into the free stream suppressing the jet into the boundary layer.

Due to the fact that initial experiments showed no level of improvement and from discussions with the project monitor few experiments were conducted on this flow making the emphasis of the research on the multiple injector optimization. It should be kept in mind, however, that only one jet position relative to the crossing shocks was conducted and generalizing these findings to all geometries is not suggested without a more detailed investigation.

Table B1. Effect of crossing shocks on the mixing characteristics of transverse jet injection

Model	Mixing Area [mm ²]	Penetration [mm]	CFOM
Without Crossing Shocks	469.98	15.32	1.414
With Crossing Shocks	351.60	14.87	1.226

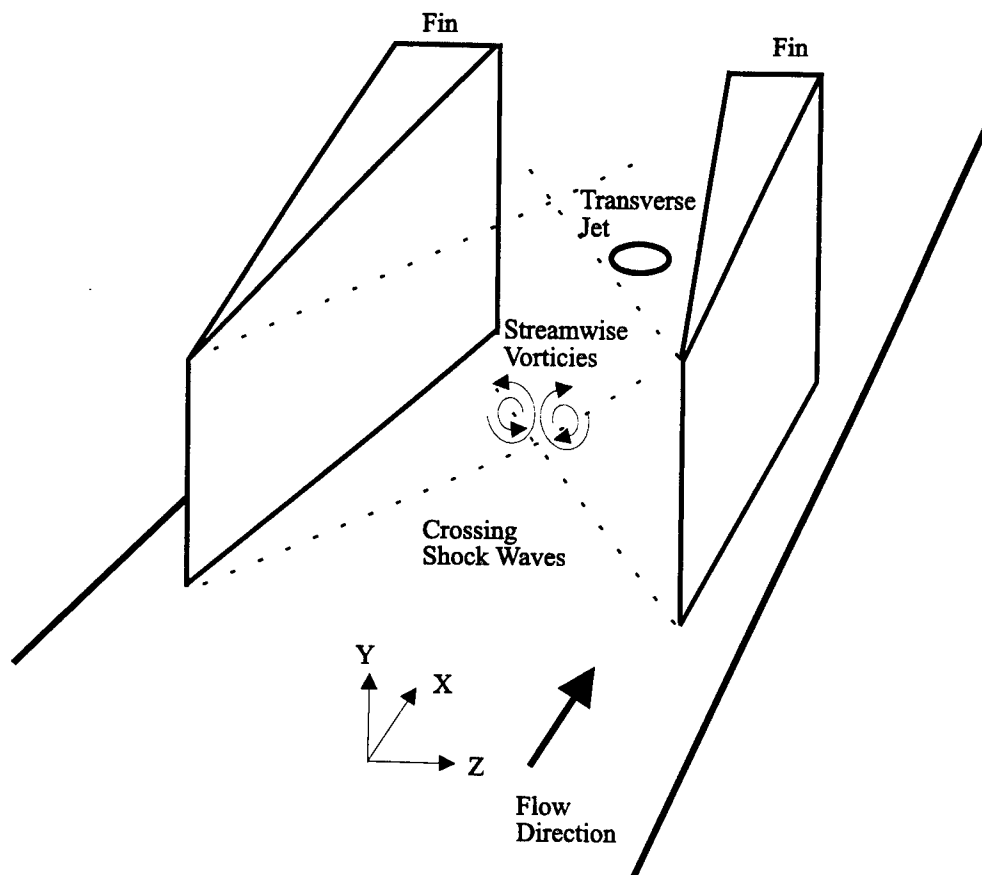


Figure B1. Schematic of a transverse jet in a supersonic cross flow with crossing-shock interaction.

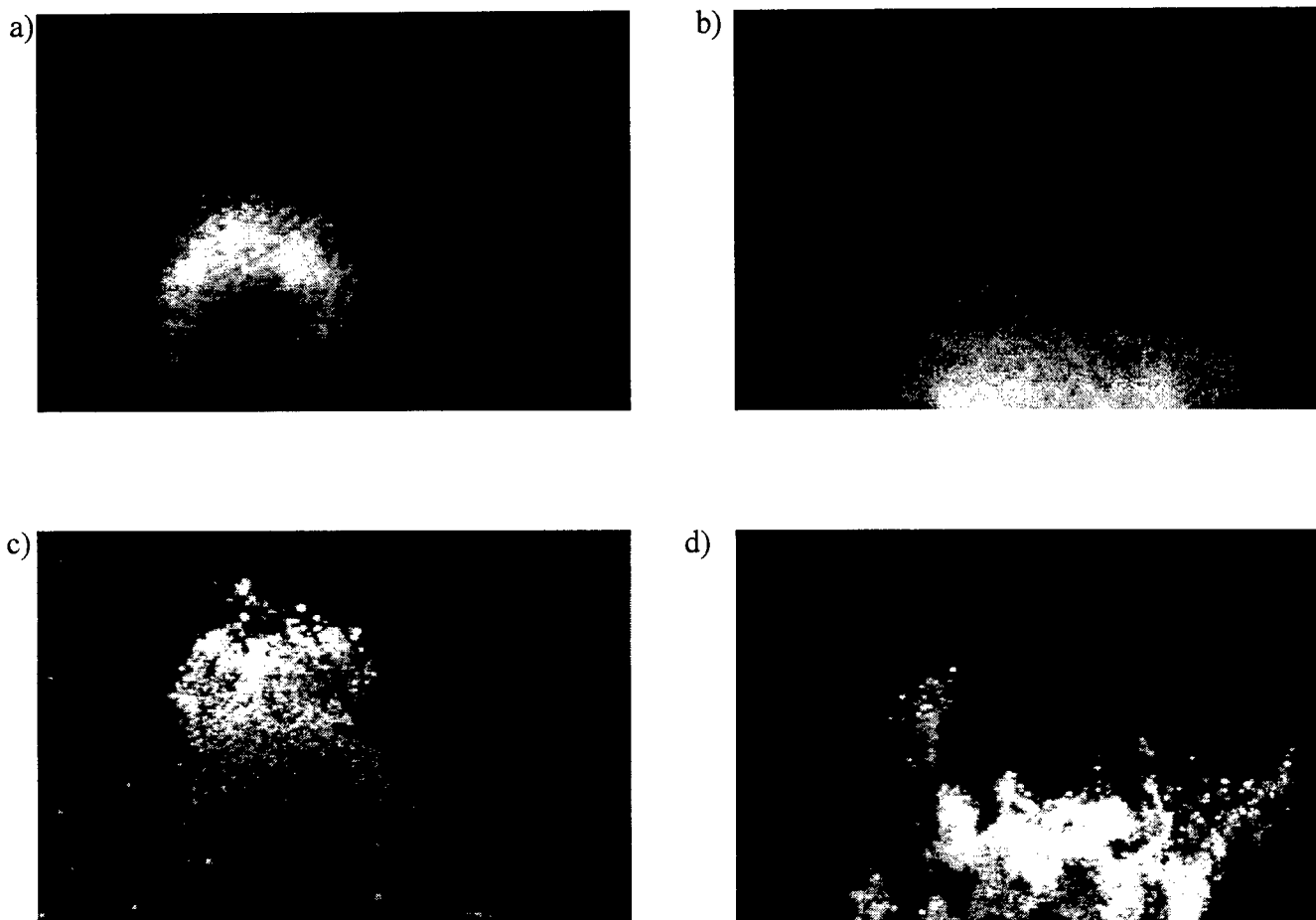


Figure B2. Spanwise view images for the averaged and instantaneous single injector without crossing shocks (a,b) and (c,d), and averaged and instantaneous single injector with crossing shocks.

Development of Multiple Metal-Dielectric Layers for 3-D MMIC

Dr. Altan M. Ferendeci
Professor
Department of Electrical and Computer Engineering

University of Cincinnati
Cincinnati, OH 45221-0030

Final Report for:
Summer Research Extension Program
Wright Patterson Air Force Base

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base
Washington, D.C.

and

Wright Patterson AFB

December 1997

DEVELOPMENT OF MULTIPLE METAL-DIELECTRIC LAYERS FOR 3-D MMIC

Altan M. Ferendeci

Department of Electrical and Computer Engineering

and Computer Science

University of Cincinnati

INTRODUCTION

At present, solid state power sources can not replace the high power slow-wave microwave tubes. But by incorporating novel power combining techniques, the overall combined power of the solid state sources may approach that of a single tube output. This is especially true for phased array antenna systems where each radiator has its own power amplifier and the total radiated power is the combined power of all the radiators. In a phased array antenna system, the transmit/receive (T/R) modules are placed on a line (linear array), on a plane (planar array) or on a circular area (circular array). In all cases, the separation distance between the antennas should be less than half the free space wavelength of the radiated signal to eliminate the main side lobes.

At microwave and millimeter wave frequencies, planar antennas are replacing the conventional antenna structures. The dimensions of the patch antennas are generally less than the required free space half-wavelength. They can be configured to be used in any array configuration. Unfortunately, if a compact module including all the electronics is to be integrated with the patch antenna, it is impossible to place all the individual circuit modules on the same substrate and at the same time satisfy the antenna spacing requirements. This becomes more important especially at millimeter wavelengths. Even though most of the system components are in the form of monolithic integrated circuits, they still occupy a large surface area of a substrate. Even if the complete circuit can be laid out on the same GaAs or InP substrate, it will be too expensive to use this type of large area substrate behind each of the radiator. Even if lumped elements are replaced by distributed elements at higher frequencies, the overall T/R module may still require a large substrate area. The complete transmit/receive (T/R) module also contains a high power amplifier. Unless precautions are taken to remove the generated heat, placing this circuit on the same substrate with other temperature sensitive circuits such as low noise preamplifiers may not be desirable.

At present, all circuit modules are placed in an enclosed casing. The power is coupled to the antenna through one end of the casing. Unfortunately, this increases the overall depth of the T/R module relative to the lateral dimensions of the antenna. These modular units have the distinct disadvantage of having a surface area to depth ratio very small. Even with these limitations, they can still be utilized in a linear or a planar array. They can not be distributed over a general curved surface. For a module to be universally conformal to any surface, the depth of the module should be negligibly small compared to the maximum lateral dimensions of the modular unit.

It is possible to place the circuit modules vertically one over the other so that the final T/R module will occupy less space. 3D module configuration dictates the development of a new packaging technology so that various circuit modules at various levels can be vertically interconnected with each other with little or no loss in signal amplitude, with no cross coupling between the elements and with manageable processing steps.

There is already considerable research and development effort being extended to incorporate vertical interconnects in the implementation of monolithic microwave integrated circuits (MMIC)[1,2]. There are two major concentrated research areas that utilize 3D vertical interconnects. First area is the use of vertical interconnects for compacting the size of an integrated circuit module [3]. In this scheme, thin multiple dielectric layers (2.5 μm) are used for connecting the various elements of the circuit together. The second major concentrated effort is to stack various circuit modules of the overall T/R module one over the other using vertical interconnects. Actually both of these concepts complement each other since the final T/R module may contain various vertically interconnected MMICs.

In either case, coupling of microwave or millimeter wave signals between the layers become a major area of research. Already, there is considerable research work going on related to the coupling mechanisms between transmission lines of the various layers [4]. Novel transmission lines are also proposed and analyzed [5-9].

In this proposal, development of a new technology to implement vertically interconnected 3D T/R module for a conformal phased array antenna system is proposed. Details of the proposed development based on deposition of multiple dielectric and conductor layers including an integral patch antenna will be given. The extensions necessary for major components of the T/R module and the specific areas that have to be developed will be discussed.

PHASED ARRAY ANTENNA MODULE

Typical front end of a T/R module for a radar system is shown Figure 1. Figure 2 shows a 2x2 modular phased array antenna system. The control and processing electronics are not shown in these figures. The lowest substrate layer contains the power amplifier. Heat can be easily transferred away from the substrate by placing the overall module on a highly heat conducting metallic surface.

Various layers of the T/R module are processed by deposition of multiple dielectric materials of known thickness followed by metal depositions immersed between the dielectric layers. As the module is processed, necessary interconnecting posts (or vias) are electro-deposited and the circuit for that layer is finally processed.

The upper layer of the module has a patch antenna which could be either rectangular, circular or ring in shape [10]. The thickness of the upper dielectric layer below the patch can be adjusted to provide the maximum radiation condition for a specific antenna [11].

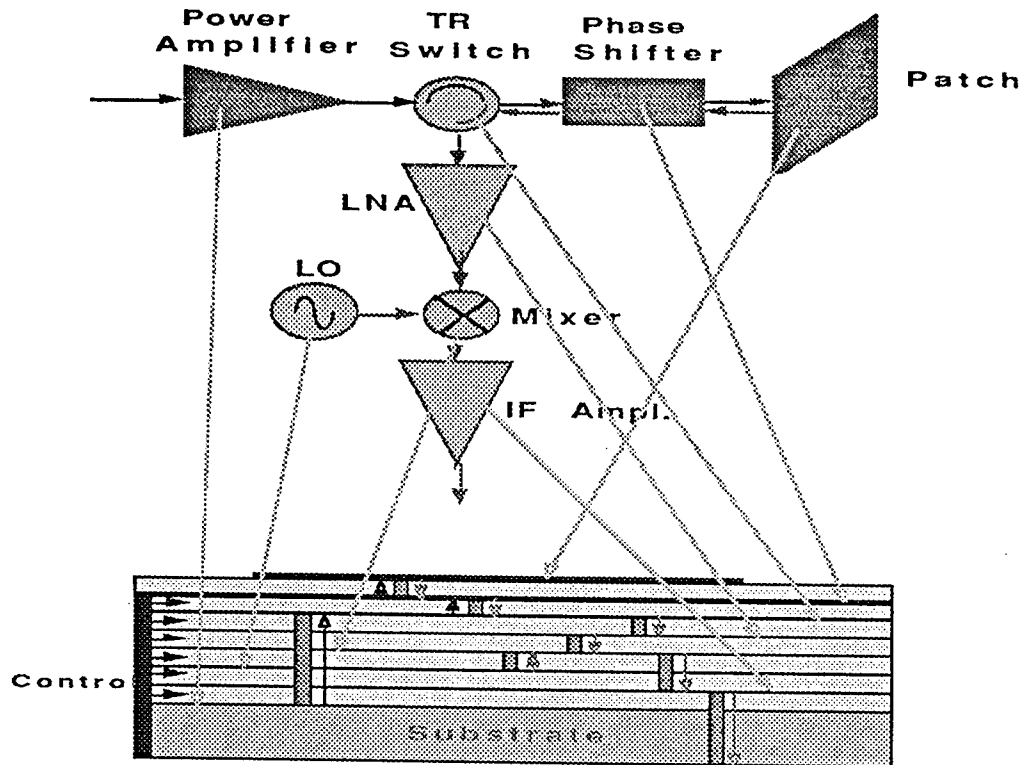


Figure 1. Major components of a T/R module.

Typical dimensions for patch antennas operating at 10 GHz and with a dielectric material of $\epsilon_r=2.5$ are given in Table I. The overall dimensions of these patches are smaller than $\lambda_o/2=15$ mm and thus can be used in phased array antenna systems since they easily satisfy the phased array antenna spacing requirements.

Table I. Dimensions of a patch antennas

($f=10$ GHz, $\lambda_o/2 = 15$ mm, $\epsilon_r=2.5$)

Patch	Dimensions (mm)	Directivity (dB)
Rectangular	$w=11.3$, $L=10.8$	6.9
Circular	$D = 10.8$	7.1
Ring	$b=4.1$, $a=2.6$	< 7.0

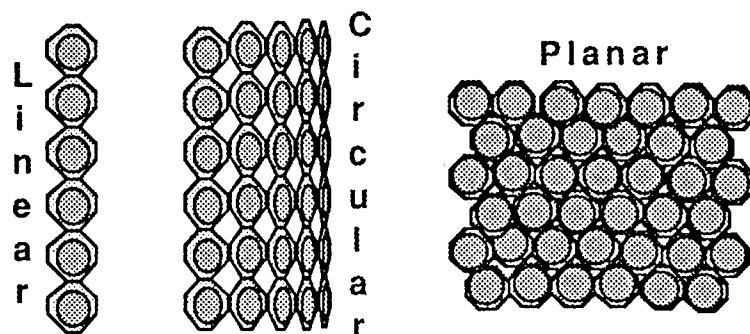


Figure 3. Distribution of patch antennas over a line, plane or a cylinder.

Figure 3 shows possible distribution of vertically interconnected 3D T/R modules over various surface configurations. These surfaces can actually be of any shape provided the radius of curvature of the surface is not too small compared to the lateral dimensions of the T/R module. Since the overall thickness of the proposed vertically interconnected 3D T/R module is expected to be a few millimeters, they can easily be distributed over any conformal surface topology. These properties may lead to vast number of applications in communication and radar systems.

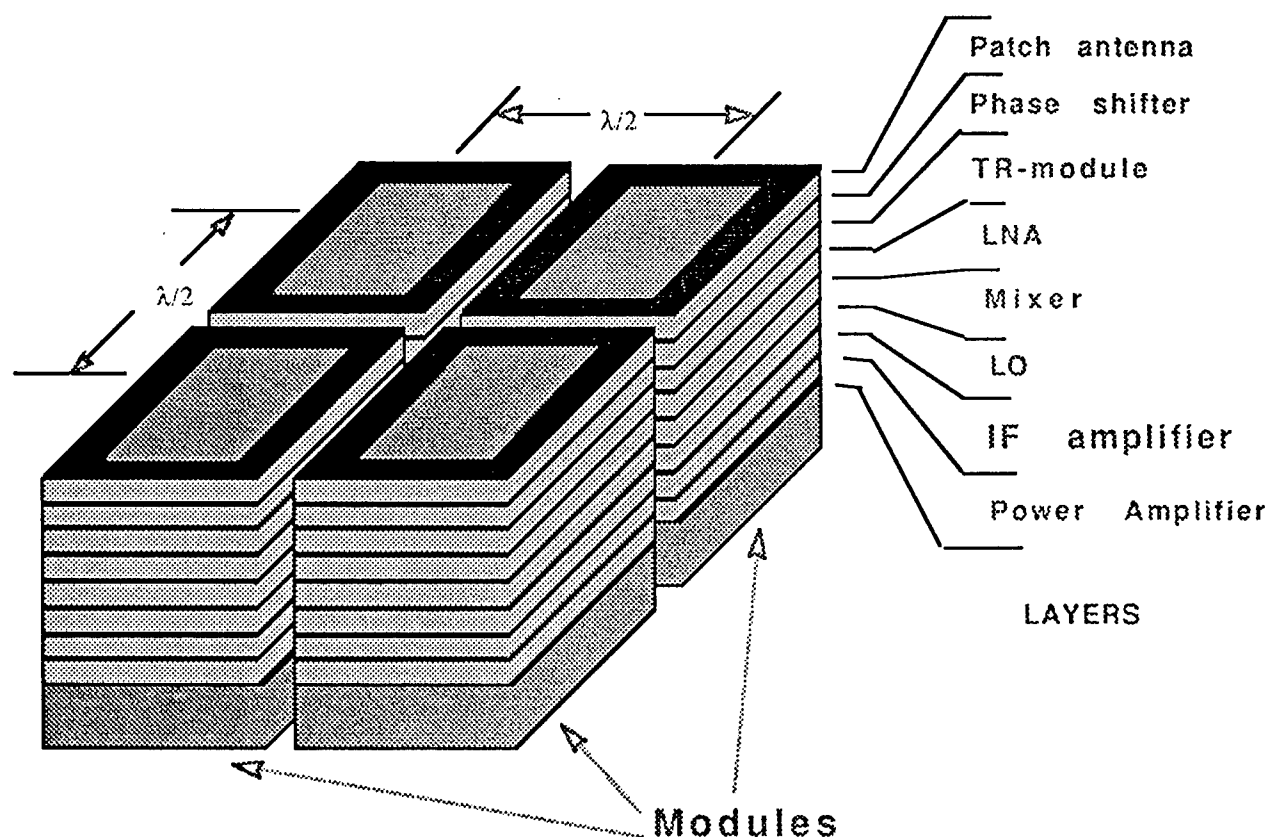


Figure 2.4 element phased array antenna system with 3D T/R modules.

Figure 4 shows the essential components of a typical representative vertically interconnected 3D T/R module. A substrate (preferably a SI-Silicon or Alumina) is the base of the building block. Since the substrate can be attached to another heat conducting material such as a metal, the power amplifier circuitry is processed on this substrate.

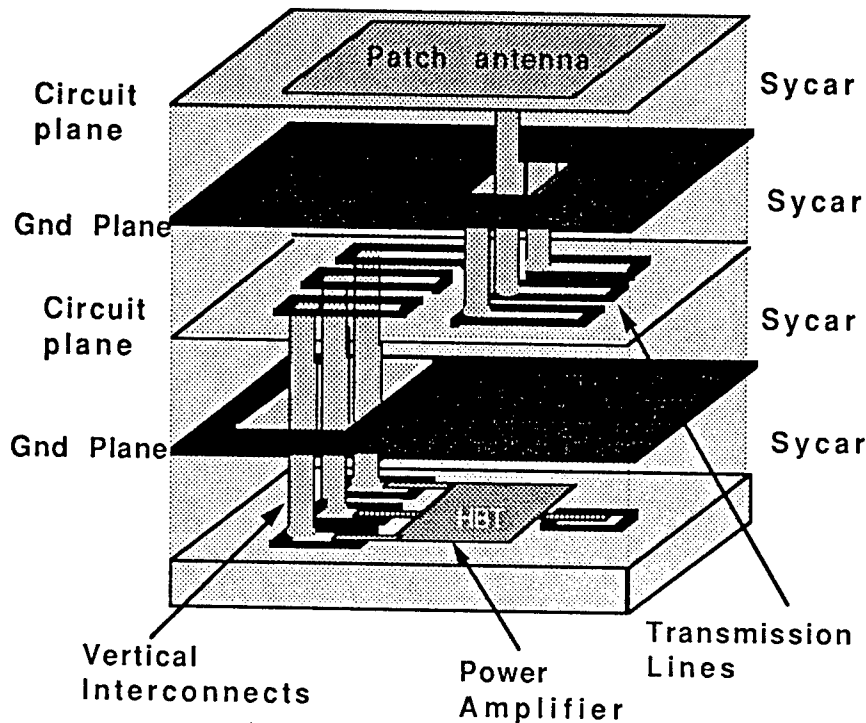


Figure 4. Representative components of a 3D module.

In addition to already developed circuit modules such as LNA, Mixer, TR switch, so on, there are four major areas that have to be further developed in implementing a vertically interconnected 3D T/R module. These are:

- a) Deposition and processing of dielectric layers
- b) Low loss vertical interconnects
- c) Transmission lines and passive circuit components
- d) Power amplifier

POWER AMPLIFIER

There are many possibilities for the implementation of the power amplifier. The complete power amplifier can be processed as a monolithic integrated circuit using either HBTs or FETs, Since there is a very successful development program on thermally shunted HBT power devices (Figure 5) at the Devices Branch of the Electronics Division of the Avionics Laboratory (AADD) of the Wright Laboratory [12], it will be an excellent opportunity to incorporate these devices in the

final 3D T/R modules. This will also allow close collaboration between the various branches of the Electronics Division. The combined expertise of the various branches will lead to the development of the best circuit topology for maximum operational efficiency.

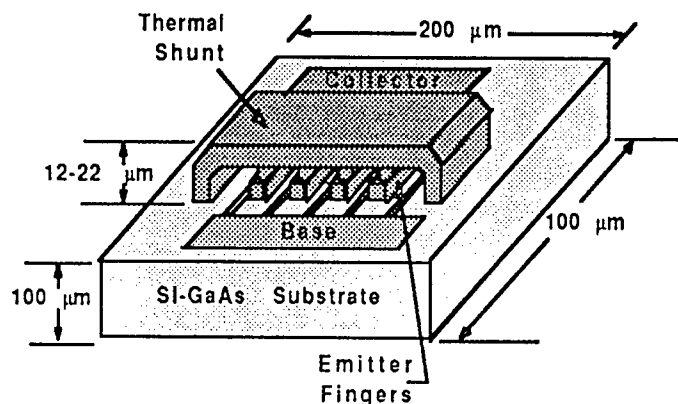


Figure 5. Thermally shunted power HBT developed at AADD of Avionics Laboratory (Ref.14).

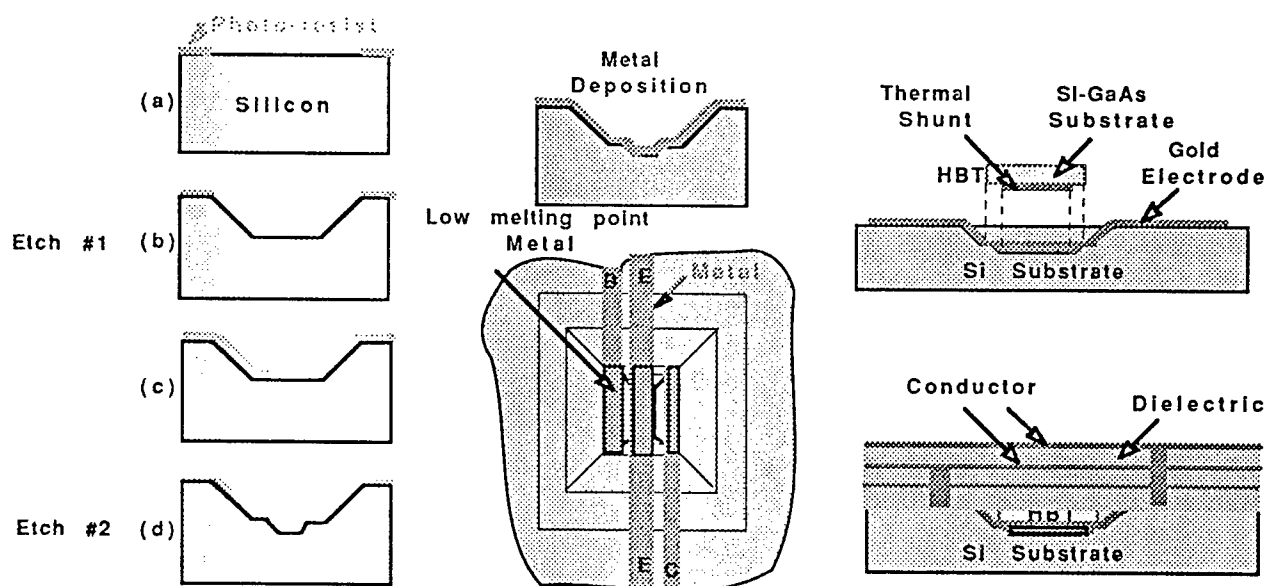


Figure 6. Micromachining steps in Silicon for attaching the HBT chip. The transistor shunt metal makes direct contact with Si-substrate.

As shown in Figure 6, the power amplifier chip is attached to the Si-substrate by micromachining an opening on the Silicon surface and flip bonding of the chip into the opening [13,14]. Since the dielectric thickness of the vertical layers are relatively thin and their thickness

may be comparable to the overall height of a thermally shunted HBT, it is recommended that the micromachining depth should be as deep as the transistor height so that the ensuing dielectric layers can be deposited uniformly over the substrate material

TRANSMISSION LINES

There are many transmission line configurations that can be used in the implementation of the vertically interconnected 3D T/R modules [15]. Microstrip lines, strip lines, slot lines and coplanar waveguides are the best possible candidates that can be used with these modules. Because of the unsymmetric nature of the slot line, it has limited applicability and in general it is used for line couplings between various layers. In many cases, circuit planes will be sandwiched between two conducting ground planes. Therefore microstrip lines are replaced by stripline type transmission lines. If the upper dielectric layer of a circuit plane is relatively thick compared to the lower dielectric layer, microstrip like transmission lines can still be implemented with a slight modification to the line characteristic. In this case, the strip is now embedded completely in a dielectric material of ϵ_r .

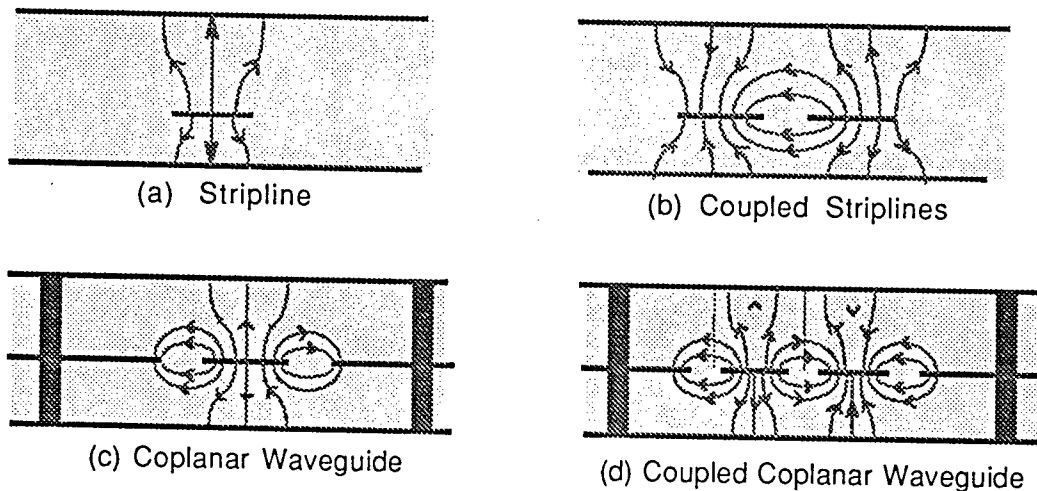


Figure 7. Transmission line suitable for 3D interconnects.

In selecting transmission line structures best suited for the proposed 3D T/R module, the lines should have low loss, easy to process and above of all they should be less sensitive to the thickness of the processed dielectric layers. If the thickness of the dielectric layer can be controlled accurately, the choice of the transmission line structure will not be critical.

For a circuit plane sandwiched between two ground planes, striplines (SL) and coplanar waveguides (CPW) are the two best transmission line configurations that can be used in implementing the distributed elements of a vertically interconnected 3D module.

Figure 7 shows a stripline and a coplanar waveguide in the presence of both upper and lower conducting planes. If the upper conducting ground plane is not present, the stripline becomes a modified microstrip transmission line imbedded completely in a dielectric material. Already successful power coupling configurations from one layer to the other are implemented when the lines are embedded within multiple dielectric layers [2].

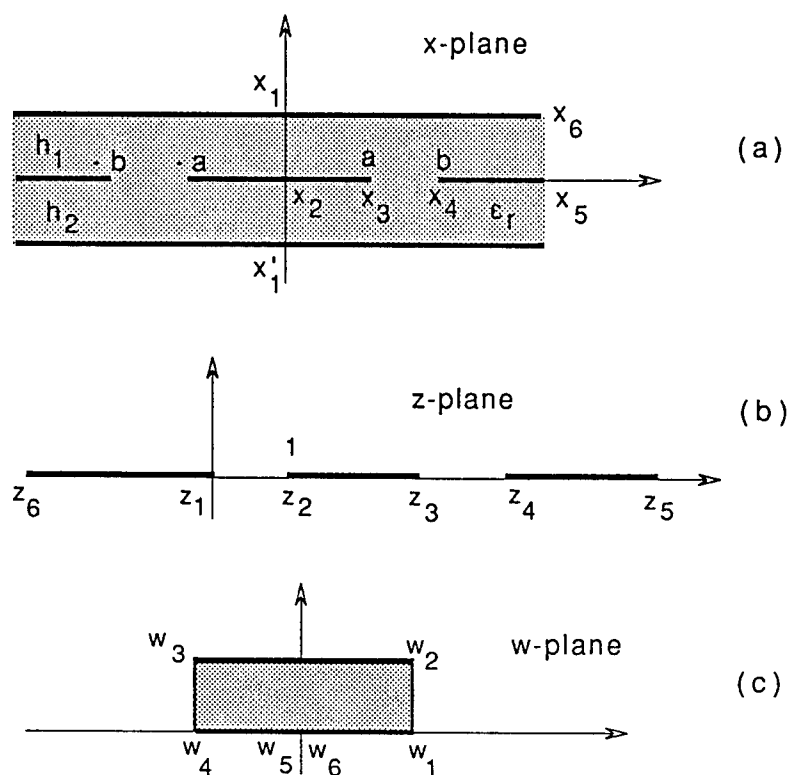


Figure 8. Conformal mapping of the CPW with top and bottom ground planes immersed completely in a dielectric material.

In order to verify the suitability of CPW for the proposed 3D modules, the effect of the upper and lower ground planes on the impedance characteristic of the coplanar waveguide is investigated. The geometry given in Figure 8a represents a typical coplanar waveguide imbedded in a dielectric material sandwiched between two conducting planes separated by distances h_1 and h_2 . A quasi-electrostatic solution for the CPW can be found by using conformal transformations [16-20]. The geometry given in Figure 8a is treated as two separate problems, one for the upper and

one for lower parts of the transmission line. Both parts can be transformed into the z -plane through the transformation

$$z = \tanh\left(\frac{\pi x}{2h}\right)$$

The transformed z -plane coordinates are shown in Figure 8b. The z -plane is then transformed into the w -plane using elliptic transformations (Figure 8c). The resulting characteristic impedance can be written as

$$Z_o = 60 \pi \frac{1}{\sqrt{\epsilon_{eff}} \left(\frac{K(k_2)}{K(k'_2)} + \frac{K(k_3)}{K(k'_3)} \right)} (\Omega) \quad (1)$$

Here

$$k_3 = \tanh(\pi a / 2h_1) / \tanh(\pi b / 2h_1)$$

$$k_2 = \tanh(\pi a / 2h_2) / \tanh(\pi b / 2h_2)$$

and $K(k)$ is the Complete Elliptic Integral of the First Kind with argument k and $k' = \sqrt{1 - k^2}$. Since the whole line is imbedded in a dielectric material, the $\epsilon_{eff} = \epsilon_r$.

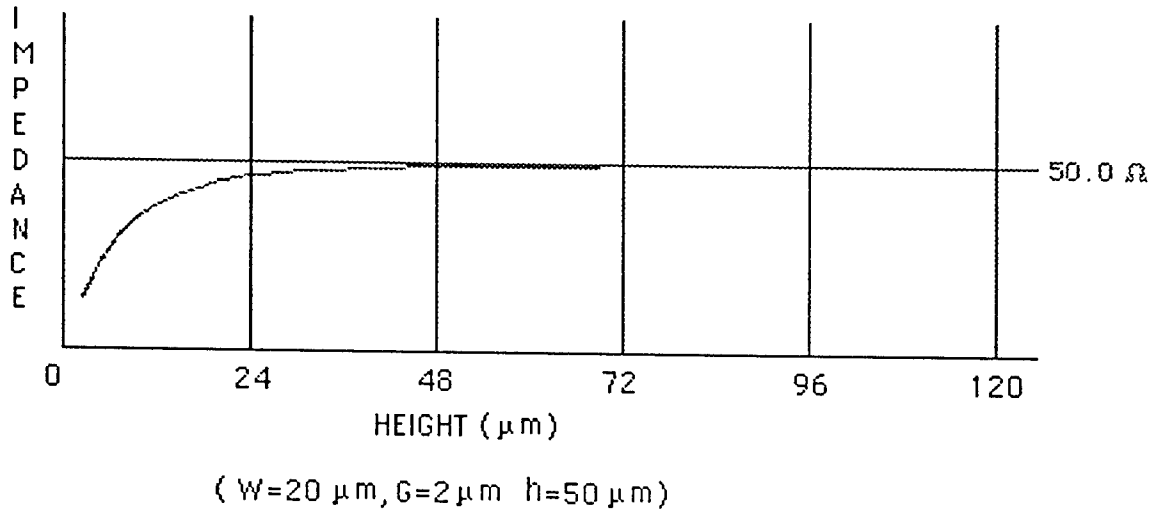


Figure 9. Impedance variation as a function of ground plane to CPW distance h for $\epsilon_r = 2.5$.

A FORTRAN program is written and the thickness of the dielectric layers between the upper and lower ground planes are taken to be equal. For a gap of $x_4 - x_3 = 2 \mu\text{m}$ and width of $2x_4 = 20 \mu\text{m}$, the variation of the line impedance as a function of the dielectric thickness h is plotted as shown in Figure 9. The impedance is not effected by h as the dielectric thickness becomes greater than twice the overall separation distance of the two side ground planes. Therefore, for dielectric layer thickness greater than $2(W+2S)$, the CPW parameters will not be critically dependent on small differences in the thickness of the deposited dielectric layers.

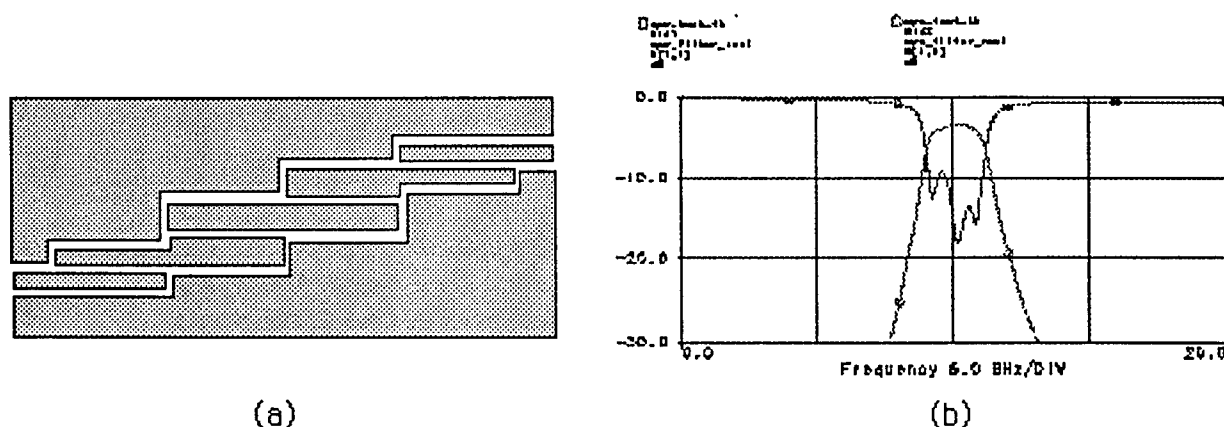


Figure 9. (a) Layout and (b) simulated S parameters of a CPW 3-section coupled equal ripple filter.

In order to show the feasibility of utilizing CPWs as distributed circuit elements, a 3 section equal ripple coupled-line filter is designed and simulated. The line parameters for the odd and even mode impedances of the coupled CPWs are calculated beginning from a low pass filter prototype using the insertion loss method. The physical dimensions of the coupled CPW lines are then found from the Line_Calc program of Libra simulation software. These are tabulated in Table II for $h=50 \mu\text{m}$ and $\epsilon_r=2.5$. The layout of the filter is shown in Figure 9a.

Table II. Parameters of the coupled coplanar waveguide filter

Section	Z_{odd}	Z_{even}	Gap (C-C)	Width	S (C-G)
1	40.2	88.9	10	292.2	24.6
2	47.3	74.2	10	506.6	118.5
3	47.3	74.2	10	506.6	118.5
4	40.2	88.9	10	292.2	24.6

Using Libra, the circuit is simulated and the resulting $|S_{21}|$ and $|S_{11}|$ are plotted in Figure 9b. It should be noted that coupled CPW line parameters used in the Libra simulations are for a CPW with the upper layer in air.

DIELECTRIC DEPOSITION.

One of the crucial components of a vertically interconnected 3D system is the choice, deposition and processing of the dielectric material. There are many different dielectric materials that are used in the processing of semiconductor devices. Table III lists the properties of some of the dielectric materials used in the implementation of 3D interconnects. The parameters in this table are for a 50 Ω microstrip transmission line. Gold metal is used to calculate the conductor losses.

Table II. Properties of dielectric materials used in 3D interconnected modules.

	BBC	Si2	Sycar	Polyimide
Deposition Process	Spin coat and bake	CVD	spin coat and bake	spin coat and bake
Dielectric Constant	2.7	4.0	2.6	3.1-3.6
Tan δ	0.004	0.03	0.003	0.001
Possible thickness (μm)	26	9	25	2.5-25
Loss/ λ_g dielectric	0.093	0.712	0.070	0.023
(dB)/cm conductive	0.812	2.307	0.838	7.327-0.876

Dielectric materials to be used in vertically interconnected microwave and millimeter wave modules should have some stringent mechanical, chemical and electrical properties. These can be listed as:

- deposition of uniform and controlled thickness layers
- adhesion to metal (gold) surface
- adhesion of metal (gold) on to the dielectric

- d) low shrinkage during curing
- e) compatibility with photoresists
- f) processable with wet and dry etching techniques
- g) maintain its properties over wide range of temperature operation.

Polyimide is one of the widely used dielectric materials in 3D interconnect circuitry. In this proposal a new dielectric material referred to as 'Sycar' developed by Hercules Inc. will be used due to its excellent moisture resistance, electrical properties, ionic purity and ease of handling. It is applied by spin coating and curing at 200° C [21].

It is very important that during the deposition process, the thickness of the dielectric layer should be precisely controlled to provide the desired thickness uniformity. Any variations in the thickness of the layers will prevent successful implementation of the circuit elements especially those of the distributed line elements whose parameters are likely to depend on the dielectric thickness. As has been discussed above, distributed line elements such as CPW are recommended for the implementation of the transmission lines since they are less dependent on the dielectric thickness provided their lateral dimensions are small compared to the thickness of the dielectric material

Mechanical properties of the dielectric material also play a very important role in the implementation of the 3D interconnects. Shrinkage during processing, stability of the material, adhesion to metal as well as adhesion of metal on the dielectric, and peeling strength of the metal over the dielectric are some of the properties that have to thoroughly investigated.

Compatibility of photoresists with the dielectric material and the type of etching process such as wet or dry etching are also major steps in the successful implementation of the 3D interconnects. It is possible that there may be a need for mechanical surface removal by lapping and polishing. Thus the dielectric material should also be robust enough to allow this type of abuse.

VERTICAL INTERCONNECTS

Another major component of a 3D T/R module is routing of microwave or millimeter waves from one level to another. In some of the 3D interconnects used in MMIC processing, novel coupling schemes between various layers are already used. If a complete T/R module as shown in Figure 1 is to be implemented, several ground planes will separate various circuit planes from each other. Therefore, it will be necessary to couple signal from one layer to the other by means of vertical posts (vias). IF CPWs are used, it is possible that parallel transmission line TEM modes may also be excited in addition to CPW TEM modes. To prevent these unwanted modes, shorting

connections between the top and lower ground planes and the side ground planes of the CPW have to be made. These will require additional vertical posts.

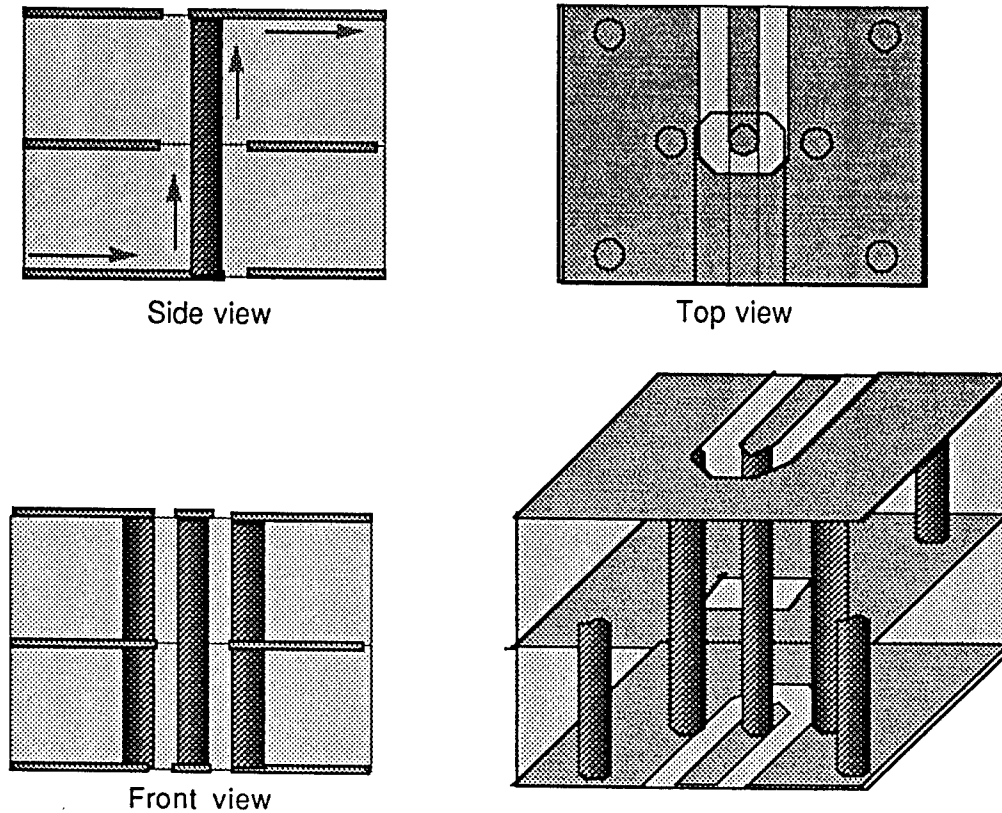
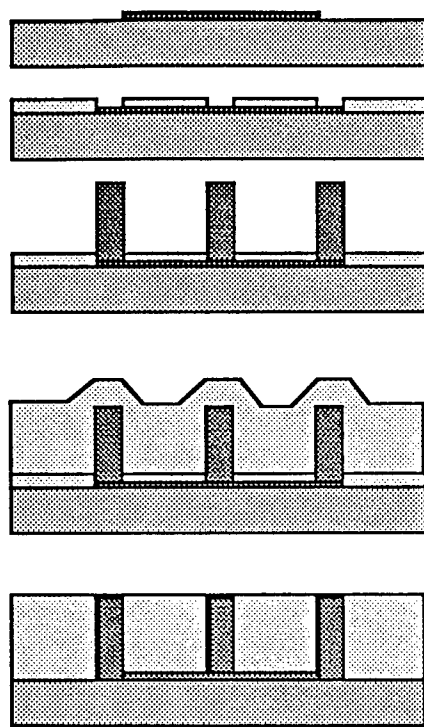


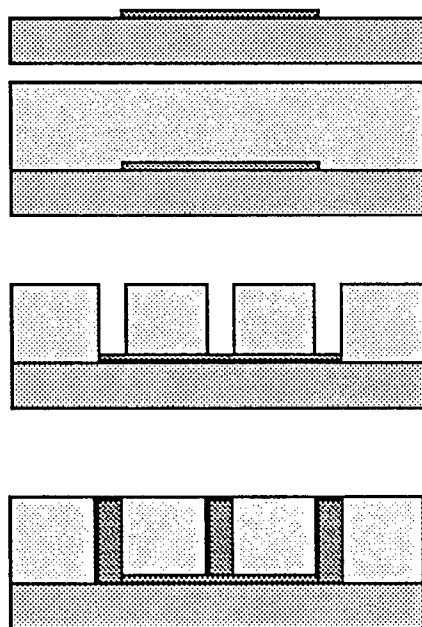
Figure 10. Vertical Interconnecting posts between layers. Additional posts to prevent possible parallel plate TEM modes are also shown.

Figure 10 shows the connecting posts (vias) from one layer to the next. Even though the posts are shown as solid uniform cylinders, their shape may be intentionally modified to decrease the return loss. Also, the cross section of the posts may not be uniform vertically as a result of the processing steps in the electrodeposition of the posts. In either case, extensive simulations and refined processing steps will be necessary to find the best possible coupling configuration. In Figure 10, additional posts connecting the ground planes and the side ground electrodes of the CPWs are also shown.



- a) Deposit metal layer for transmission line or ground plane and process
- b) Deposit a thin dielectric layer
- c) Open via (post) holes
- d) Electrodeposit metal posts to required height
- e) Spin coat and process dielectric
- f) Mechanically planarize surface

(a)



- a) Deposit metal layer for transmission line or ground plane and process
- b) Deposit dielectric layer to the required thickness
- c) Open via (post) holes in dielectric
- d) Electrodeposit metal posts to required height

(b)

Figure 11. Two alternate processes for vertical posts.

Figure 11 shows two possible multi layer processing steps for realizing dielectric and vertical post deposition. The process shown in Figure 11a is easier to implement and is expected to generate slightly non-uniform vertical posts. This process involves first the deposition of a

transmission line or a ground plane. A thin dielectric layer is then deposited over the metal. Using photoresist, the post locations and sizes are exposed. Metal posts are then electrodeposited to the required height. The photo resist is then removed and a thick dielectric layer is spin coated and cured. This process will form mesas over the posts. Mechanical lapping may be required to planarize the upper surface of the dielectric material to the depth of the metal posts. Even though this process seems to be attractive in the overall implementation of the 3D interconnects, required mechanical lapping may not be desirable.

The second process shown in Figure 11b begins with metal deposition for the transmission line or a ground plane. The dielectric material is then spin coated and cured to the required final thickness. Photoresist is then used to locate the post locations. Through a wet or dry etching process, the dielectric material is removed from these openings for via holes. Gold metal is electrodeposited in these holes to the height of the dielectric layer. This process requires anisotropic etching of vertically uniform via holes in a thick dielectric material. If this can be achieved, uniform cross sectional electrodeposition of metal posts will be possible.

MODELING AND SIMULATIONS

In order to implement an optimum vertically interconnected 3D T/R module, extensive component and circuit simulations are necessary.

Even though circuit simulation software such as Libra, MDS and Compact are available, they do not include some of the modified circuit topology that will be used with the 3D T/R modules. For example, the CPW model used in Libra is for a CPW with the upper surface exposed to air. In the CPWs that will be used in the 3D modules, the electrodes will be completely immersed in a dielectric material and there will also be an upper ground plane in addition to the conventional lower ground plane. In the quasistatic solution given in Equation 1 for the CPW, both complete dielectric immersion and the presence of the upper and the lower ground planes are included. Since Libra does not have this type of circuit element in the Library, either a new library element has to be added to the simulator, or a new simulation procedure has to be developed. This is also true for the coupled CPWs. The filter simulations presented in Figure 9 are made using the conventional CPW models that are available in Libra. The response of the filter will be modified by the presence of the upper ground plane as well as complete immersion of the CPW in the dielectric medium. The lack of these elements in the present simulation software libraries will generate large simulation errors especially when resonant circuit elements are used in a circuit.

Some of the simulation software also lack detailed device modeling especially for the power devices [19,20]. Since temperature effects modify the operating characteristics of power devices

such as HBTs, new nonlinear models has to be introduced into the software or new device files have to generated and used with SPICE like programs.

There are also many electromagnetic simulators available. Among these, Microwave Lab and Sonnet are the two software which are exclusively dedicated to modeling planar circuit elements. CPWs and the vertical interconnects will be simulated using these software. Even though these programs require large computer memory to run complicated geometry, they provide both electric and magnetic field distributions, current distributions in the conductors as wells the S parameter data at a given frequency. These are essential in the successful implementation of various circuit components especially in finalizing the shapes and locations of the interconnects for minimum return loss.

It can be concluded that successful implementation of 3D T/R modules will require extensive and careful device and circuit modeling and simulations in addition to developing novel processing techniques.

SHORT TERM GOALS

The short term goals of this research proposal is to develop technology for vertically interconnected MMIC. The major emphasis in this phase of this project is to develop a reliable and controllable dielectric deposition process so that necessary number of dielectric layers can be deposited.

The initial phase of the program will be concentrated in the deposition of the dielectric and subsequent metal layers. The uniformity as well as the reproducibility of the dielectric layers are very important. Thickness control and adhesion to various surfaces such as Silicon and a metal surface are major parts of the development program. If a successful dielectric deposition process is developed and the thickness of the deposited layers can be controlled to within a micrometer, the ensuing circuit designs will be relatively simplified.

WORK SCOPE:

A key technology will be developed for 3D MMIC. Emphasis will be placed on the deposition of uniform and controllable dielectric layer thickness for vertical interconnects. Various processing technologies will be developed for uniform via holes and associated metal posts. Coplanar waveguides and other transmission lines sandwiched between layers will be implemented.

Work Period: 1997 (1 year)

Total Funding: (\$ 25 K, detailed budget is attached)

Man Power: 1 professor and 1 graduate student

The detailed worksteps are as follows:

- Sycar will be initially used as a dielectric material. Other possible dielectric materials such as polyimide and BCB will be investigated.

- Sycar multi-layer process will be developed with emphasis on

 - thickness control

 - uniformity of thickness

 - compatibility with wet and dry etching techniques

- Multi-layer vertical interconnect technology will be developed with emphasis on

 - via holes

 - gold interconnect posts

 - metal dielectric interface

- Extensive study of the best combination of waveguide structures for vertically interconnected multi-layer circuit formation will be studied. Emphasis will be concentrated on Coplanar-Waveguide because of less dependence on thickness control. Microstrip, stripline, et al. will be studied provided uniform thickness control is achieved during the dielectric deposition process.

- Passive components (capacitor and inductor) will be developed.

- Electromagnetic simulation of vertically interconnected transmission lines between layers will be extensively studied using existing software (MicroWave Lab) and in-house developed numerical techniques.

LONG TERM GOALS

Long term goals of the project will be to implement complete T/R modules and test them in a 2x2 array for their applicability in a conformal phased array antenna system. This challenging task will be achieved through extensive theoretical analysis, simulations and experimental testing of various components and circuit modules.

PROGRAM SCHEDULE

Month	1	2	3	4	5	6	7	8	9	10	11	12
Process development												
-Dielectric Layer												
Material Study												
-Layer Deposition												
-Metalization												
-Via Process												
Passive component												
-design												
-fab and test												
Electromagnetic and circuit												
Simulation												
-multi-layer design												
-coupling study												

References:

- [1] I.Toyoda & T. Tokumitsu, "Miniaturized Microwave Circuits on Multilayer MMICs," MWE '93 Microwave Workshop Digest, 335-38(1993).
- [2] T.Tokumitsu, K.Nishikawa, K. Kamogawa, I. Toyoda & M. Aikawa," Three Dimensional MMIC Technology for Multifunction Integration and Possible Applications to Masterslice MMIC," IEEE MTT-S Digest, MMW (1996).
- [3] I.Toyoda, T. Tokumitsu & M. Aikawa, "Highly Integrated Three-Dimensional MMIC Single-Chip Receiver and Transmitter," IEEE MTT-S Digest (1996).
- [4] M.J.Tsai, C.Chen, T.S.Horng & N. Alexopolus, " Multiple arbitrary Shape Via Hole and Air Bride Transitions in Multilayered Structures," III MTT-S Digest, 707-10 (9916).
- [5] K.Atsuki & K.Li, "Characteristics and Analysis Methods of Transmission Lines for Multilayer Substrates," MWE '93, Microwave Workshop Digest, 323-26(1993).
- [6] H. Mandai, T.Tsuru & M. Kato, "Characteristic of Ceramic Multilayer Substrates and applications for RF Circuits," MWE 93 Microwave Workshop Digest, 327-30(1993).
- [7] H.Yabuki & S.Maeda, "Characteristics of Multi-layer Soft Substrates and their applications to RF Circuit Modules," MWE '93 Microwave Workshop Digest, 331-34(1993).
- [8] T. Krems, W. Haydl, H. Massler & J. Rudiger, "Millimeter Wave Performance of Chip Interconnections Using Wire Bonding and Flip Chip," IEEE MTT-S Digest (1996).

- [9] P.C. Hsu & C.Nguyen, "New Multilayer Planar Transmission Lines for Microwave and Millimeter Wave Integrate Circuits," IEEE Trans. MTT-43, 1809-13(1995).
- [10] Giudong Guo, "Active Phased Array Antenna Systems Using Ring Resonator Oscillators," M.S. Thesis, ECECS Dept. University of Cincinnati (June 1995).
- [11] K. Kamogawa, T. Tokumitsu & M.Aikawa, "A Novel Microstrip Antenna Using Alumina-ceramic/Polyimide Multilayer Dielectric Substrate," IEEE MTT-S Digest (1996).
- [13] P.M.Zavrocky, T.Earles, N.L.Pakrovsky, J.A.Gree & B. Burns, "Fabrication of Vertical Sidewalls by Anisotropic Etching of Silicon (100) Wafers," J.Electrochem. Soc., Vol. 141, 3182-88(1994).
- [14] J.M.Kim, W.N.Carr, R.J.Zeto & L. Poli, "Reactive Ion Etching Techniques for Silicon Sidewall Angle Control in Micromachining," J. Electrochem. Soc., 1700-05(1992).
- [12] R.Dettmer, T.Jenkins, J.Barrette, C.Bozada, G.DeSalvo, J.Ebel, J.Gillespie, C.Havasy, C.Ito, C.Pettiford, T.Quach, J.Sewell, D.Via & R. Anholt, "Effect of Device layout on the Thermal Resistance of High Power Thermally-Shunted Heterojunction Bipolar Transistors,' IEEE MTT-S Digest (1996).
- [15] R.Sturdivant, C.Quan & J. Wooldridge, "Transitions and Interconnects Using Coplanar Waveguide and Other Three conductor Transmission Lines,' IEEE MTT-S Digest, (1996).
- [16] C.P.Chen, "Coplanar Waveguide: A Surface Strip Transmission Line Suitable for Nonreciprocal Gyromagnetic Device Applications"" IEEE Trans MTT-17, 1087-90(1969).
- [17] G. Ghhione & C. U. Naldi, "Coplanar Waveguides for MMIC Applications: Effect of Upper Shielding, Conductor backing, Finite-Extent Ground Planes and Line to Line Coupling", IEEE Trans MTT-35, 260-67 (1987).
- [18] S.S. Bedair, M.I. Sobhy, "Accurate formulas for Computer-aided design of shielded microstrip circuits." IEE Proc. Vol.127, Pt.H, 305-308 (1980).
- [19] R. Kulke & I.Wolff, "Design of Passive Coplanar Filters in V-band," IEEE MTT-S Digest (1996).
- [20] C. Veyres & V. F. Hanna, " Extension of the application of conformal mapping techniques to coplanar lines with finite dimensions," Int. J. Electr., Vol.48, 47-56(1980).
- [21] J.K.Bard & R.L.Brady, " A new Moisture Resistant Liquid Encapsulant," IEEE 42nd ECTC Conference Proceedings, 1028-22(1992).

DEVELOPMENT OF A PROTOTYPE TO TEST AND DEMONSTRATE THE
MANUFACTURING-ORIENTED, DESIGN-DIRECTED COST ESTIMATION (MODDCE) FRAMEWORK

Allen G. Greenwood
Associate Professor
Department of Industrial Engineering

Mississippi State University
P. O. Box 9542
125 McCain Engineering Building
Mississippi State, MS 39762

Final Report for:
Summer Research Extension Program
Wright Laboratory / Manufacturing Technology Directorate
Manufacturing and Engineering Systems Division

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Department of Industrial Engineering and the Office of the Vice President for Research
Mississippi State University

February 1998

DEVELOPMENT OF A PROTOTYPE TO TEST AND DEMONSTRATE THE
MANUFACTURING-ORIENTED, DESIGN-DIRECTED COST ESTIMATION (MODDCE) FRAMEWORK

Allen G. Greenwood
Associate Professor
Department of Industrial Engineering
Mississippi State University

Abstract

This report discusses the development and application of a software prototype that demonstrates the application of the MODDCE (Manufacturing-Oriented, Design-Directed Cost Estimation) framework and provides a proof of concept for using an object-oriented approach to implement the framework. The MODDCE framework defines the modeling and analysis progression from product features to an enterprise view of the production of a product. The most design-relevant portion of the framework is implemented through an object-based representation of an integrated product/process cost assessment environment. This representation is based upon the premise that cost assessment must concurrently consider three primary design dimensions, or what may be referred to as the product triad -- form, material, and process. The prototype is referred to as the OCEANS (Object-based Cost Assessment and Estimation for design Support) system and its initial focus is on the design of aircraft structures. The OCEANS prototype is a critical first step towards a comprehensive software system and application environment that effectively integrates cost assessment into the design process. In addition, the prototype provides an important means to identify, discuss, and address a variety of cost/design issues with industry IPTs (Integrated Product Teams) and product designers.

DEVELOPMENT OF A PROTOTYPE TO TEST AND DEMONSTRATE THE MANUFACTURING-ORIENTED, DESIGN-DIRECTED COST ESTIMATION (MODDCE) FRAMEWORK

Allen G. Greenwood

Introduction

In order to develop affordable products, designers and Integrated Product Teams (IPTs) must place cost/manufacturability assessment, along with performance, at the core of the Integrated Product Process Development (IPPD) or concurrent engineering process. Designers and IPTs need to effectively address both product and process issues early in a product's design; i.e., they must have tools available that are sensitive to process parameters as well as product parameters. Cost, schedule, and risk must be assessed during design in manner similar to the way aerodynamics, weight, etc. are assessed. Therefore, there is a strong need for advanced design technologies to effectively support the transformation of customer requirements into a product by providing cost/manufacturability measures of the design and concurrently considering the production process in conjunction with the product's intended form and material.

Unfortunately, such cost/manufacturability assessment tools are not readily available today. A review by Thomas [1] of papers published in major journals in the past five years, resulted in the following conclusions: (1) research in manufacturing cost estimation has involved rather few people and as a result there are no broadly applicable mathematical models, (2) there is widespread dissatisfaction with the way overhead or indirect costs are handled, and (3) the lack of broadly applicable cost models leads to few efforts directed towards their integration into CAD/IPPD systems. Greenwood [2] adds the following issues to the list of concerns with the current state of manufacturing cost estimating within the IPPD process: (1) Cost model drivers need to be more closely related to the characteristics of manufacturing processes and operations. The process needs to be designed as much as the product; all too often the product is designed and then the process (manufacturing) must be designed to accommodate it or the product must be redesigned to accommodate the process. (2) Cost model drivers need to be more closely related to product features. (3) Manufacturing process design and resulting costs need to be considered earlier in the design of the product. (4) Product and process design must be sensitive to the existing or planned manufacturing state, especially in an agile environment; i.e. a design may change based on current factory utilization. Cost models need to capture the dynamics and stochastic nature of manufacturing. (5) The relationship between product/process costs and indirect costs need to be better defined and understood. Indirect costs, both manufacturing and non-manufacturing, need to be "modeled" so that they can be addressed during design. (6) Legacy cost system data are often not tracked at the level of detail needed in design. (7) Legacy cost system data are often based on processes and methods that are no longer used.

In order to design affordability into products, manufacturing and cost considerations must be addressed early in the design process. Since this ability is lacking in today's design environment, it is important

to freshly examine how costing *should* be done in order to support design and new costing methodologies must themselves be *designed* to fit into tomorrow's environment. The prototype described in this report is a step towards meeting these needs. It is entitled OCEANS (Object-based Cost Estimation and Assessment for design Support). The prototype application is based on the most design-relevant portion of the Manufacturing-Oriented, Design-Directed Cost Estimating (MODDCE) framework [2, 3]. The framework defines the modeling and analysis progression from product features to an enterprise view of the production of a product. The OCEANS prototype is also based on an object-based approach [4] that was developed to provide a structure for integrating product and process considerations early in the design phase of a product's life cycle. Both the MODDCE framework and the object-based structure are defined below.

The MODDCE Framework

The OCEANS prototype is based upon the Manufacturing-Oriented, Design-Directed Cost Estimation (MODDCE) framework [2, 3]. This framework defines the modeling and analysis progression from product features to an enterprise view of the production of a product. As illustrated in Figure 1, it is composed of five levels that progress from a hierarchical decomposition of a product's features through three intermediate levels to an enterprise model that links product features with manufacturing and non-manufacturing entities throughout the enterprise. The highest or enterprise level provides the most comprehensive analysis capabilities for assessing the effect of changes in product and process design on the enterprise and the resulting effect of those changes on "total" cost. As shown in Figure 1, each level is linked to a set of design parameters that drive the types of costs that are considered at that level. The dashed-line box in the left-hand portion of Figure 1 is the portion of the framework where a product's design evolves, in terms of what it looks like, what it is made of, and how it is constructed. Hence, the focus of this research and development effort is within this design-relevant portion of the framework.

The progression from one level of analysis to another in Figure 1 is accomplished in stages; the numbered arrow lines in Figure 1 represent stages. For example, in order to move from the hierarchical decomposition of a product's features to the corresponding manufacturing *elements* that create those features, a stage (Stage 1 in Figure 1) maps or associates features with transformation elements. The manufacturing elements may be a specific manufacturing process or a general process family.

The second stage combines manufacturing elements into *operations*. Operations define specific equipment, labor skill requirements, and tooling and consider precedence among elements and operations. The operations level provides the first, yet limited, opportunity to make product/process cost trades.

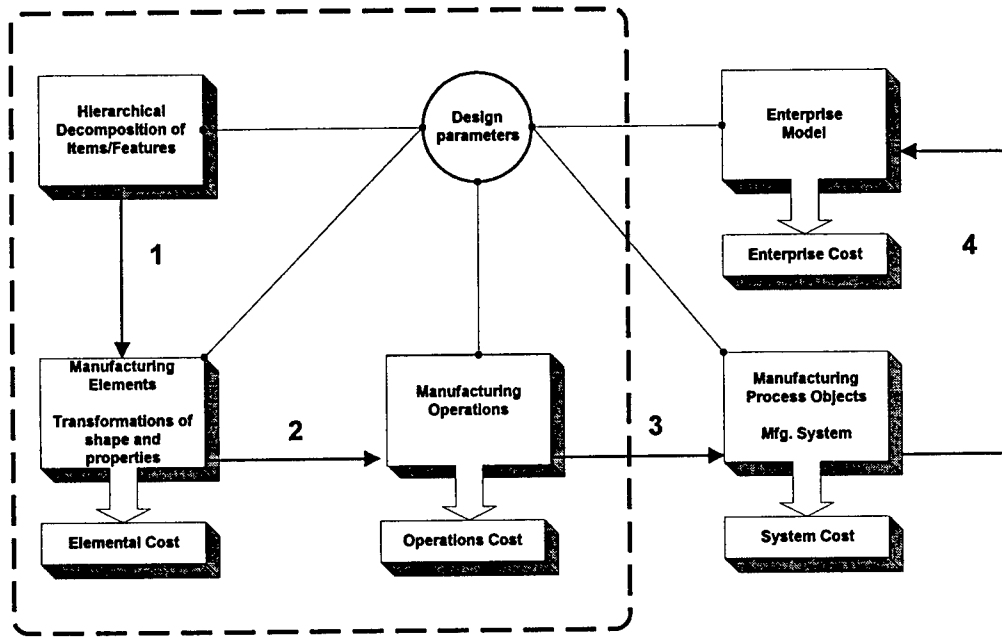


Figure 1: MODDCE framework provides the basis for “designing” and integrating costing methodologies that link product features to an enterprise view of the production of the product.

The third stage transforms discrete process operations into system-level manufacturing process *objects*. It is at this level, and through the use of the objects, that manufacturing process operations are placed in the context of an actual production environment; i.e., the product is integrated into the fabric of a manufacturing environment. As a result, the objects become the basis for the design of production operations and facilities by considering such things as product movement, storage and handling issues, resource availability, interference with other products, grouping of elements into cells, etc. Process objects are dynamic (time sensitive) and stochastic (the ability to incorporate random events) – both of which are common characteristics of manufacturing systems, but are seldom considered in the product design phase. It is only through the dynamic and stochastic capabilities of the manufacturing process object that a system’s performance can realistically be evaluated and virtual manufacturing can truly be achieved. It is only through manufacturing process objects and the interaction among objects that manufacturing performance, including costs, can be evaluated. System-level analysis provides a means to identify bottlenecks and other potential problems before the product is committed to production. It is only at this level that the impact of a product’s characteristics on a production process can be assessed and evaluated. This is especially important in an agile environment or for accommodating small batches of dissimilar products (e.g. items for the defense industry being produced on commercial lines). In fact, the design of the product may change based on the current state of the manufacturing facility, e.g. available resources, supplier constraints, etc.

The fourth stage moves modeling and analysis to the fifth and highest level -- the enterprise level. Enterprise-level costing extends design sensitivity beyond direct manufacturing costs to include non-

manufacturing and indirect costs. The impact of changes in product/process design on indirect entities in the enterprise is assessed, and conversely, the impact of changes in indirect entities on manufacturing processes, and hence on the product, are also assessed. For example, specification of a new material during product design will not only affect operations in terms of the type of equipment used, labor skills, tooling, scrap rate, etc. but may also affect numerous indirect entities in the organization, including training, purchasing, warehousing, etc. Normally these effects are assumed to be captured in the application of very general indirect or overhead rates. However, enterprise-level costing permits a more realistic assessment of the impacts on indirect activities. In order to realize enterprise-level cost estimation, the firm's business processes must be identified, modeled, and linked to the appropriate manufacturing process object(s). In this case, enterprise is defined in its broadest context and includes all non-manufacturing entities in the company (administration, purchasing, etc.), as well as external entities such as customers and suppliers.

The Object-based Foundation

Object-based technologies were investigated, as part of another research project [4], and found to offer a strong promise for facilitating the development and use of an integrated costing environment in support of concurrent product and process design. Object orientation provides a more natural and easier means -- as compared to more traditional systems analysis, design, programming, and database approaches -- to represent and model complex systems from individual components. In object-oriented applications, data objects that represent real-world entities are often the focus; whereas, applications that employ a conventional programming approach are often procedure focused. Therefore, from a modeling perspective, as noted by Khoshafian and Abnous [5], decomposition of a system from an object-oriented perspective is more intuitive and involves well-defined behaviors and encapsulated algorithms; in contrast, procedure-based decomposition is unnatural, in most cases.

The OCEANS prototype is based upon the object-based representation of an integrated product/process cost assessment environment that was developed in [4]. This representation is based upon the premise that cost assessment must concurrently consider three primary design dimensions, or what may be referred to as the product triad -- *form*, *material*, and *process*. That is, in order to adequately cost an item, one must know what it looks like, what it is made of, and how it is produced. The last element is the one that has received the least consideration in the past. However, in order to design and build affordable products and to make meaningful trades among alternatives, all three elements must be an integral part of the design process and "cost" is a natural metric to tie them together. While the amount and level of information available early in the design phase may be limited, manufacturing process still must be addressed and assumptions must be made as to how the product will be produced.

The object-based, product-triad approach is represented conceptually by the triangular structure in the lower portion of Figure 2. The three primary design dimensions, or product triad -- form, material, and process -- are defined as object hierarchies and are represented as circles in the figure. Secondary objects, shown as the

intersection of two primary objects and represented by ovals in the object structure are complex objects, in that they inherit properties from two primary objects. For example, the Form-Material object inherits properties from both the Form and Material hierarchies. There is a second type of complex object, shown as the intersection of all three primary objects and represented by the triangle in the center of the object structure. This Form-Material-Process (F-M-P) object inherits properties from all three hierarchies. This object is especially important since it provides the basis for cost assessment; i.e. costing cannot be completed until the design has considered the item's form, material, and manufacturing process.

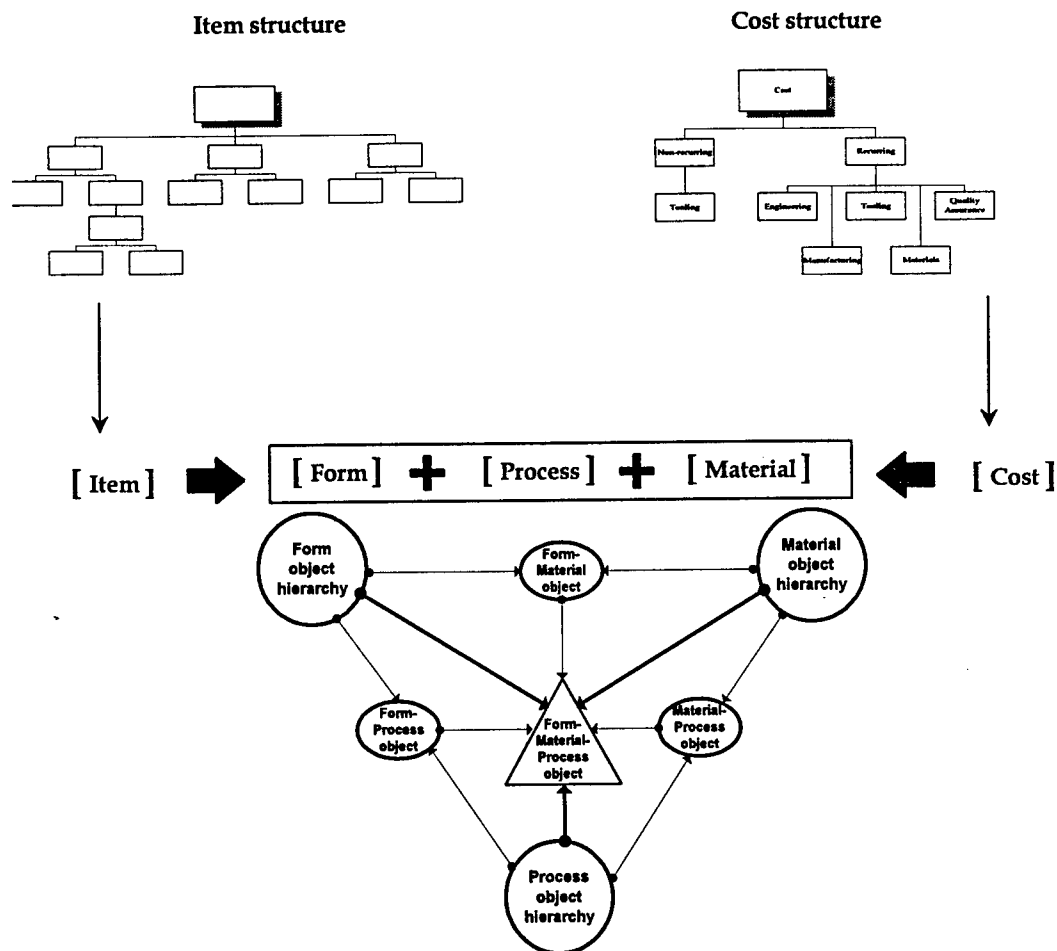


Figure 2: Cost assessment based on a Form-Material-Process instance.

Two additional elements are required in order to complete a cost assessment. First, the item being designed must be considered in its context as a part of some larger system or product, as illustrated by the item structure in the upper left portion of Figure 2. The item structure primarily provides programmatic information, such as quantity, schedule, etc. These properties are often important factors in design since, for example, low production rate or quantity may preclude an item from being produced by certain processes because they would

result in prohibitive costs and become unaffordable. Cost structure is the second element that must be combined with the Form-Material-Process design definition. The cost structure as shown in the upper right portion of Figure 2, defines the types of cost that are addressed in the analysis.

Approach to Prototype Development

The OCEANS prototype is a critical first step towards a comprehensive software system and application environment that effectively integrates cost assessment into the design process. It was developed in order to demonstrate the application of the MODDCE framework and provide a proof of concept for using an object-based approach in order to implement the framework. In addition, the prototype was developed in order to provide an important means to identify, discuss, and address a variety of cost/design issues with industry IPTs and designers. The demonstration and application of this tool in an industrial environment will result in a better understanding of the types of data, knowledge, models, interfaces, etc. that are required in order to represent, process, and display information in a manner that fosters the concurrent consideration of product/process cost during design.

The OCEANS system is a fully-functional software application; e.g., it is built upon a sound database structure, employs an intuitive interface, provides meaningful real-time feedback to the user on the cost impact of design changes, and provides an efficient means for maintaining the data. Of course, being a prototype, the system, is limited in the breath and depth of its application domain. However, these limitations are by design, since it would be premature to fully develop an application until many of the critical issues are better understood, modeled, and represented.

The data used to populate the prototype is a small portion of the information provided in the multi-volume *Manufacturing Cost/Design Guide (MC/DG)* that was developed by Battelle, as part of the Air Force's Integrated Computer Aided Manufacturing (ICAM) program [6, 7].

A portion of this research involved an investigation into the type of software that should be used to prototype and develop the application. Due to the growing dominance of the Microsoft *Windows* operating system and the plethora of development languages and applications that it supports, it was decided to develop the OCEANS system in a *Windows 95/NT* environment.

A variety of off-the-shelf applications were reviewed for inclusion in OCEANS. Several potentially attractive applications were strongly considered. Knowledge Base Engineering's (KBE) *Integrated Knowledge Environment (IKE)* possessed some of the capabilities needed to develop OCEANS. However, KBE was developing another product, *Object cZar (OZ)*, which would provide many more of the capabilities required by OCEANS. For example *OZ* provides a means to easily incorporate flexible form, material, and process hierarchies, as well as cost and item structures. In addition, *OZ* provides an effective means to create and manage "collections" of objects. Therefore, it was decided that it would be more effective in the long run to wait for a more mature release of *OZ*. Incorporation of technologies from *OZ*, expected during the first quarter of 1998, will greatly enhance the next version of OCEANS.

Cognition Corporation's *Cost Advantage (CA)* provides strong manufacturing process cost modeling capabilities. While *Cost Advantage* is a mature product in the *UNIX* environment, introduction of the *Windows* version has been continually postponed. However, its *Windows NT* version seems certain to be released in the first quarter of 1998. Therefore, the next version of OCEANS will most certainly be greatly enhanced by utilizing the many capabilities of *Cost Advantage*. A completely new Cognition product, *Knowledge Center*, is expected to be released in early 1998 and, based on preliminary information, it will enhance the OCEANS system's object structure as well as its data/knowledge representation and processing.

Since none of the above applications were available in time to develop the prototype, the initial version of the OCEANS system was developed in *Visual C++* with Microsoft *Access* providing the relational database capabilities. However, the next version of OCEANS will rely less on *C++* and more on the off-the-shelf capabilities of *OZ*, *Cost Advantage*, and *Knowledge Center*.

User "view" of the oceans system

The description of the OCEANS prototype is provided in terms of two "views" of the application. The primary or user view is the environment in which the designer or IPT interacts with the system. The secondary or administration view is concerned with updating existing information and entering new information into the system. Screen shots from each of the views are used to describe the capabilities of the OCEANS system.

The basic interface screen for the user view is shown in Figure 3. The user view contains six user input "tabs" – material, form, process, DICE, Assembly, and draft image -- that are used to specify a design for costing purposes. The "draft image" tab in Figure 3 shows a general picture of the product that is being designed. In this case, a fuselage panel is being designed. It is composed of three basic items –skin panel, stringer(s), and frame(s).

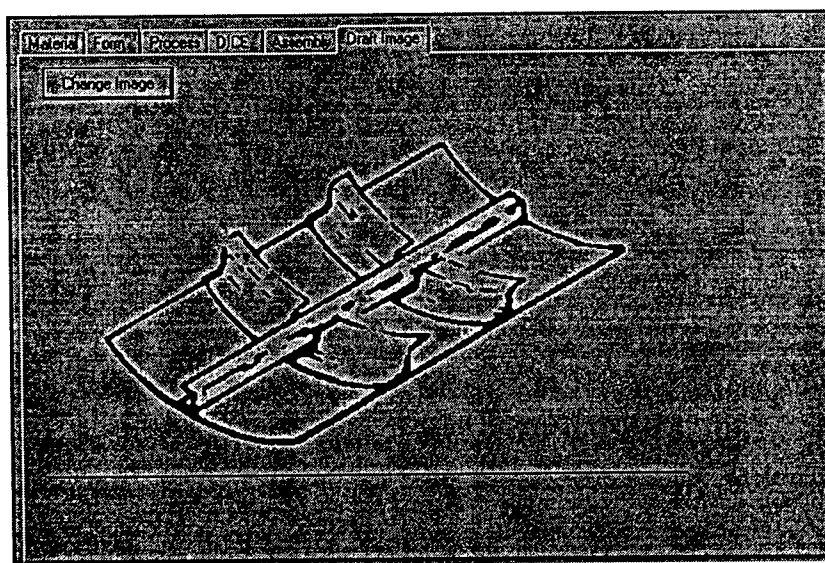


Figure 3 – Basic user view showing the product that is being designed.

The three primary tabs in the OCEANS system – Material, Form, and Process – correspond to the “product triad” that was described earlier in the “The Object-based Foundation” section. This follows the premise that an item’s cost cannot be estimated until it has been determined what the item looks like, what it is made of, and how it will be produced. It should be noted that the properties that are displayed or controlled on each tab represent a minimal set of properties to cost an item. These properties are based on the cost estimating relationships provided in [6, 7]. This initial version of OCEANS is based on a literal implementation of the *MC/GD*. Future versions would employ a flexible structure that would incorporate a variety of costing approaches and data.

Figure 4 illustrates the “Material” tab. Two elements of the design are specified through this screen – the items that comprise the product (and their quantity) and the type of material for each item. In all cases throughout this example, white boxes on the screen denote areas where the user can enter or change a value, such as the specification of material Titanium 6Al-4V3 in Figure 4. Grayed-out panels in the screen display information that cannot be changed through the particular tab that is being displayed, such as the price of the material in Figure 4. In this case, price is not a variable that the designer can change – it is done separately through the administration view described later. However, the process specification, even though it is grayed out in Figure 4, can be changed by the designer, but through the Process tab.

The item list for the product that is being designed is controlled through the Material tab screen. In the example in Figure 4, the product is composed of four items – “skin,” “stringer,” “Frame, Zee,” and “Frame, Angle.” Since “Frame, Angle” is highlighted in the list in Figure 4, all of the displays will relate to this item. The number of items that are used in the product is specified to the right of the item name; in this case, there is one “Frame, Angle” used in the design. Also note that the material that is specified in Figure 4 is actually for the item “Frame, Angle.” Items are easily added to and deleted from the items list through a button control that is a part of the general toolbar. (The toolbar controls are not shown in the screen shots; they are similar to standard toolbars found in all *Windows* applications.)

The graphs in the right panel of Figure 4 display the relationship between recurring and non-recurring labor hours and the primary cost driver, in this case the length of the item. The dot on the line graph indicates where this item’s cost driver falls in the relevant range of the relationship; in this case the item’s length is three feet and at the lower end of the estimating range. The item’s length is specified through the Form tab that is described below. The cost estimating relationships that are plotted in the graphs in Figure 4 are obtained from the *Manufacturing Cost/Design Guide (MC/DG)* [6, 7]. In this case, CED-T-1B refers to the specific cost estimating data for producing a cylindrically-contoured angle member from titanium using the Brake/Hot Stretch process.

Material Form Process DICE Assembly Draft Image

Item Setup

Item Name: Frame, Angle X: 1

skin
stringer
Frame, Zee
Frame, Angle

Form: Angle, Cylindrically Contoured

Process: Brake/Hot Stretch

Basic Material Property

Material: Titanium-6Al-4V3

Price: 1700 \$/50 Ft

Description: Titanium-6Al-4V3(N/A)

Labour (M/H): 0.72082153
Length: 3.00 12
Recurring Cost

Labour (M/H): 20215.0000
Length: 3.00 12
Non-recurring Cost

Figure 4 – Basic user view for specifying the items that comprise the product and their materials.

Material Form Process DICE Assembly Draft Image

Item Name: Frame, Angle X: 1 Material: Titanium-6Al-4V3 Process: Brake/Hot Stretch

Basic Property of Form

Form:

☒ Angle, Cylindrically Contoured
☒ Angle, Straight Member
☒ Channel, Contoured Member
☒ Cylindrically Curvature Skin
☒ Lipped-Hat, Noncylindrically
☒ Lipped-Zee, Noncylindrically

Length: 3.00 feet
Width: 0.42 feet
Thickness: 0.04 inch
Bend Radius: 2.00 inch

☐ Symmetric ☐ Uniform

Labour (M/H): 0.72082153
Length: 3.00 12
Recurring Cost

Labour (M/H): 20215.0000
Length: 3.00 12
Non-recurring Cost

Figure 5 – Basic user view for specifying the form of the product.

Figure 5 provides an example "Form" tab. In this case, the "Frame, Angle" item has an angle cross-section and is cylindrically contoured, as shown by the selected (highlighted) form property in Figure 5. In addition to specifying shape, the Form tab is used to enter an item's basic dimensional properties, such as length, width, and thickness. The next version of OCEANS will include a hierarchy of forms. The use of an object-based hierarchical approach will permit each form to have its own set of properties, as well as inherit properties from its "parent" in the hierarchy. It is anticipated that group technology will facilitate the classification of candidate forms.

The Process tab shown in Figure 6 is used to specify the primary manufacturing process for producing an item, in this case, "Brake/Hot Stretch." The OCEANS system's process-related properties include labor rates and learning curve factors. The next version of OCEANS will include a hierarchy of processes and the list of candidate processes will include only those that are compatible with the specified form and process properties.

Implementation of the assembly process, under the OCEANS system's current database structure, presented some development complications. The problem appears to be solved but since the approach has not been fully tested, it is not included in this version. This may be a moot point because the utilization of a more object-oriented database structure, e.g. through *OZ* or the *Knowledge Center*, should eliminate the problem. However, the screen layout for the Assembly tab is provided in Figure 7 in order to illustrate its intended operation. All of the items included in a design would appear in the lower right window, thereby providing a pick list of items to be assembled. Selected items are moved to the window in the lower left portion of the screen. The method for fastening the items, along with other requirements and parameters, are specified in the boxes in the upper left portion of the screen.

Cost Reports

Figure 8 provides an example of one of the two types of cost reports that are available from the OCEANS system. This summary view reports only the top-level costs -- total recurring cost, total non-recurring cost, and average cost per shipset -- for the entire product that is being designed. In this case, the costs are for the entire fuselage panel that was specified and described above; it is named "Sample1." This view displays aggregate costs for each design alternative that is being considered (i.e., each alternative that is open on the desktop) on a separate line. The view window is also referred to as the Project Desktop since it provides access to project-level variables, such as design quantity and the constant-year dollars for the project. A toggle control button on the main toolbar controls whether the detailed cost report is displayed or hidden.

A detailed cost report is prepared for each alternative design; i.e. a cost report is "attached" to each user view that was described above. A toggle control button on the main toolbar controls whether the detailed cost report is displayed or hidden. A portion of the detailed cost report is shown in Figure 9. Detailed cost information is provided for each item in the design and includes not only a further breakdown of the cost elements (direct labor, material, etc.) but the key cost factors that are used in the cost calculations (learning curve factor, labor rates, burden rates, etc.).

Material Form Process **DICE** Assembly DraftImage

Item Name: X 1 Material: Form:

Process Property

Process: Labor Rate: Dollar/MH

Learning Curve Factor: 2

DICE: B

Recurring Cost

Non-recurring Cost

Figure 6 – Basic user view for specifying the primary manufacturing process.

Material Form Process **DICE** Assembly DraftImage

Assembly Property

Process: Assembly Name:

Fastening Method:

Installation Requirement:

Perimeter: Inch Num. of Fasteners: Inch

DICE

Recurring Cost

Non-recurring Cost

Assembly Items Setup

Items selected to be Assembled:

Applicable Items:

Buttons: Add, Delete, Property

Figure 7 – Basic user view for specifying the assembly of items.

Project Operation

Name:

Design Quantity:

Year Dollar:

Cost Estimation Summary for alternatives:

Alternatives	Total RC	Total NRC	Average C...
Sample1	\$ 150,677	\$ 16,884	\$ 838

Figure 8 – Summary cost report for a quick cost comparison of alternative design.

Cost Report for: Design Quantity:

Description	# of Labour	LC	Labour Rate	Labour RC	Material	MR Cost	Parts Per AC	Prod RC	NR/La
skin	1.58 MH	1.69	\$25.00/MH	\$66.60/parts	\$504.00/p...	\$570.60/p...	1	\$ 114,121	67.8
stringer	0.50 MH	1.29	\$25.00/MH	\$16.09/parts	\$33.93/parts	\$50.02/parts	1	\$10,005	59.0
Frame, Z...	1.28 MH	1.29	\$25.00/MH	\$41.21/parts	\$42.33/parts	\$83.54/parts	1	\$16,708	221.0
Frame, A...	0.82 MH	1.13	\$30.00/MH	\$27.79/parts	\$21.42/parts	\$49.21/parts	1	\$9,843	215.0

Total Recurring Cost: US\$

Total Non-recurring Cost: US\$

Average Cost Per AD: US\$/AC

Figure 9 – Detailed cost report for a specific alternative.

Both cost reports -- summary and detailed -- are updated as soon as a change is made in the design specification, thereby providing immediate feedback to the designer or IPT on the affect of the change.

Administration "view" of the OCEANS system

The second view offered by the OCEANS system is referred to as the administration view. It provides the capability to maintain the information in the system -- update existing information (e.g. material prices, labor rates, and burden rates) and enter new information (e.g. new forms, materials, processes, and estimating relationships) into the system. The administration part of OCEANS is used to maintain information that is not under the control of the user (designer, IPT) and is usually common across all projects. It is maintained centrally; whereas, the user portion is controlled and maintained locally. Recall the user can easily create new projects and alternative designs within projects.

There are actually two modes -- browse and change -- for the administration portion of OCEANS. While both modes have the same interface, the browse mode allows users to only view the information from the administration perspective; whereas, the change mode allows users to modify the information. Typically, the primary user of the OCEANS system, designers and IPTs, would not have authority to utilize the change mode of the administration view. However, it may be useful for the user to browse or query the basic information included in the system through the administration view. Access to the information in the different administration view modes is controlled through passwords or other security measures.

Screen shots are again used in this section to illustrate the interface and describe the information administration capabilities of the OCEANS system. A "tabbed" interface is employed, similar to those described earlier in the user view section. The screen shots reflect the change-mode interface that allows information in the system to be modified, added, or deleted.

Figure 10 shows the administrative view for information related to materials. All of the materials contained in the system are listed in the window in the right-hand portion of the screen. When a material is selected (highlighted) in the window, its property values are displayed in the left-hand portion of the screen. If the user has proper authority, they may modify these values (price, units, description, etc.). In browse-only mode, these property values would be displayed with a gray rather than white background, indicating that they can not be changed. The OCEANS database is updated once one of the modify, insert, or remove buttons are selected. In the browse-only option, the modify/insert/remove buttons are "grayed-out" and hence inactive. As mentioned earlier, the next version of OCEANS will contain a more extensive, and dynamic, list of properties. This change will convert the materials list to a materials hierarchy and the properties that are displayed will depend on the materials class that is selected. For example, a heat treat property will only be present on metallic materials and not present on composites. The OCEANS system will also have the ability to connect to other databases in order to directly provide property values. For example, material prices may be maintained in

Material Form Process Classification CED CEDIndex

Edit Area

Material Label: Aluminum-2014-T6

Price: 17.00

Price Unit Label: \$/SQ.FT

Classification: Aluminum

Material Description:
Aluminum-2014-T6(N/A)

Material Table Operation

Aluminum-2014-T6
Aluminum-7075-T6
Steel-4340 180 Ksi H.T.
Titanium-6A1-4V1
Titanium-6A1-4V2
Titanium-6A1-4V3
Titanium-6A1-4V4

Modify Insert Remove

OK Cancel Apply

Figure 10 – Administration view for maintaining information on materials.

Material Form Process Classification CED CEDIndex

Edit Area

Process Label: Brake Form

Classification: Brake Form

Recurring: 25.00 US\$/MH

Non-Recurring: 30.00 US\$/MH

Learning Curve: 0.90

Process Description:

Process Table Operation

Brake Form
Brake/Hot Stretch
Brake/Roll
Creep Form
Drop Hammer
Farnham Roll
Hot Press
Preform/Hot Size
Rubber Press

Modify Insert Remove

OK Cancel Apply

Figure 11 – Administration view for maintaining information on processes.

another database; therefore, the value displayed will not be directly editable in OCEANS. In this case, the property value for material price will actually be a pointer or path designation to an element in another database.

Figure 11 shows the administration view for manufacturing processes. Its interface and operation is similar to that for the Materials tab. The only process-related properties that are contained in this version of OCEANS are recurring and non-recurring labor rates and a learning curve factor. The next version will extend this set of properties considerably.

Figures 12 and 13 show the administrative views that provide the means for defining cost estimating relationships and assigning them to specific form-material-process combinations. As discussed earlier, the cost estimating relationships are those provided in the *MC/DG* ([6, 7]). The *MC/DG* reports contain only a graph of each relationship – the underlying data or equations for the relationships are not available. For the purpose of this prototype, piecewise-linear approximations of the relationships are used. Figure 12 shows the interface for the CED tab. In this example, the relationship labeled CED-A-1A is specified. (The CED-A-1A label follows the naming convention used in the *MC/DG*.) The X,Y values that are used to define the piecewise-linear relationship is entered in the CED Data Set section of the screen. A graph of the relationship is automatically displayed in the “CED Preview” window included in the lower right-hand portion of the screen. This estimating relationship is linked or assigned to a specific form-material-process combination through the “CED Index” tab.

The “CED Index” tab is shown in Figure 13. In this example, any item that is made of “Aluminum,” is of the form “Angle, Cylindrically Contoured,” and is produced using a “Brake Form” process uses the estimating relationship CED-A-1A for recurring costs and CED-A-1B for non-recurring costs. If a form-material-process combination has not been assigned, the system will warn the user of that fact and will use the first compatible process for the specified form-material combination, if one exists. Incompatibility of specific form-material-process combinations may be defined through this window by selecting the appropriate values for form, material, and process in the top part of the screen in Figure 13 and then checking the “Incompatible” box in the lower left-hand portion of the screen.

A later release of the OCEANS system will contain a more flexible cost module that will allow a variety of types of cost relationships obtained from different sources to be defined and selected by the user.

Future Research and Development Directions

The prototype that is described in this report was demonstrated to the Lab focal point for this project, representatives from a major aircraft manufacturer, and two software development companies. Everyone was very supportive of the work to date and encouraged future development. They all felt that it is critical to have tools/environments like the OCEANS system available early in the design process in order to develop affordable products. All of the reviewers had very helpful comments and suggestions, many of which the author was aware of but inclusion of those capabilities in the prototype was beyond the scope of the project. Many of these ideas are presented below. They set an agenda for future development.

Material Form Process Classification **CED** CED Index

Edit Area:

CED Label: **CED-A-1A**

Dependent Variable Label: **Length**

Unit Label: **Feet**

CED Table Operation

CED-A-1A
CED-A-1B
CED-A-6
CED-S-7
CED-T-1A
CED-T-1B
CED-T-3A
CED-T-3B
CED-T-4
CED-T-5A
CED-T-5B

Modify
Insert
Remove

CED Data Set

Recurring Non-recurring

	X	Y		X	Y
1	2.00	0.13	1	2.00	12.50
2	3.00	0.16	2	7.00	18.75
3	4.00	0.19	3	12.00	20.00
4	5.00	0.22	4	0.00	0.00
5	5.99	0.25	5	0.00	0.00
6	6.00	0.31	6	0.00	0.00
7	12.00	0.60	7	0.00	0.00

CED PREVIEW

Recurring Cost: MH 00.130
Length: 2 2.00 12

Non-recurring Cost: MH 12.500
Length: 2 2.00 12

OK Cancel Apply

Figure 12 – Administration view for defining cost estimating relationships.

Material Form Process Classification **CED** CED Index

Pointers:

Material: **Aluminum**

Form: **Angle, Cyclindrically Contoured**

Process: **Brake Form**

CED Index

Available CED1:

☐ Incompatible
☐ CED Unavailable

CED-A-1A
CED-A-1B
CED-A-6
CED-S-7
CED-T-1A
CED-T-1B
CED-T-3A
CED-T-3B
CED-T-4
CED-T-5A

Available CED2:

CED-A-1A
CED-A-1B
CED-A-6
CED-S-7
CED-T-1A
CED-T-1B
CED-T-3A
CED-T-3B
CED-T-4
CED-T-5A

☐ Is subject to Bend Radius

Critical Point: inch

Minimum: inch

OK Cancel Apply

Figure 13 – Administration view for assigning cost estimating relationships.

As indicated earlier, while the OCEANS prototype is a fully-functional system, it has limited capability, is based on a limited application domain, and includes a portion of the data and expertise from a single source, the *Manufacturing Cost/Design Guide (MC/DG)* ([6, 7]). However, these limitations are within the defined scope of the project. While knowingly limited in terms of its breadth and depth, this prototype demonstrates a proof of concept for an object-based cost assessment environment that supports the design process. As such, it provides a very important foundation for future developments. For example, it is a vehicle for generating dialogue with industry in order to further define the capabilities that are required in order to effectively support the design/IPPD processes. Throughout this report, references have been made to how the OCEANS system will be improved in subsequent versions. Those enhancements are briefly summarized below. This is followed by a preliminary description of a more complete OCEANS application environment and how it might be utilized in the design process. The report concludes with a brief discussion of the next logical step in the development process.

The development of this prototype has demonstrated that the use of an object-based approach is not only viable but an extremely important technology on which to build a system that will effectively provide an integrated cost assessment capability during design. However, the OCEANS system must be more object-based than its current implementation. Some of the major capabilities that are needed are: more object-oriented data structures, hierarchies with property inheritance, knowledge storage and processing, flexible and open cost models, and links to other platforms, data, and programs. As mentioned earlier, these most likely will be achieved by utilizing such software as *OZ* and *Knowledge Center*. The system also needs broader and deeper information on product forms, materials, and manufacturing processes. The OCEANS system needs to incorporate a means to better link cost elements to design parameters and be able to better handle product assembly and integration issues. Many of these concerns can be resolved through interactions with industry. In fact, many aspects of the OCEANS system will be greatly enhanced by applying the system in a real-world setting and implementing the information and suggestions that result from these endeavors into future versions. Demonstrating and testing the OCEANS system in a real-world environment provides a rich opportunity to capture and represent information on real products, processes, and cost models, as well as expertise from designers and IPTs.

The OCEANS prototype is a first step towards a comprehensive software system and application environment. A vision for how such a system would effectively integrate cost assessment into the design process is illustrated in Figure 14. The system would provide a strong link between members of an IPT and the product/process design environment. It would also be considered an evolving application, in that its capabilities and resources would continually be updated and enhanced based on lessons learned by the IPTs and through research projects. The system would itself be a valuable research tool and would provide a powerful learning environment for academe and industry.

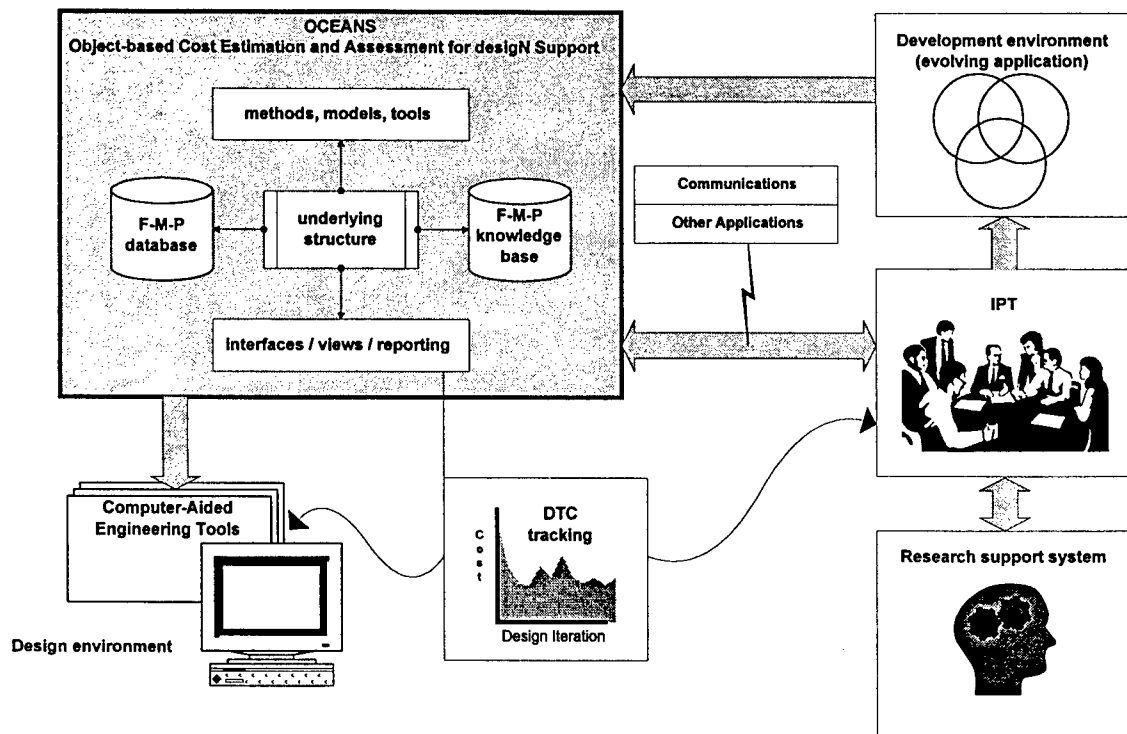


Figure 14: The envisioned application environment for the OCEANS system.

The primary components of the system are shown in the shaded OCEANS box in Figure 14. The F-M-P (Form-Material-Process) database contains instances of the object classes and is populated with a variety of design alternatives so that new design considerations are able to draw upon and effectively use this history. The knowledge base contains rules, element descriptions, notes and annotations from past use, etc. that further define and support the development of relationships among the objects that are used to assess product/process cost. The system is composed of numerous methods, models, tools, etc. that define the behavior of objects and provide the computational capability of the system. Providing effective feedback to designers and IPTs, in forms that are meaningful and useful to them, is critical to any system's success. Therefore, "interfaces" is defined as a primary element of the system. Interfaces include the means for all users of the system to effectively and efficiently provide input to and receive output from OCEANS. Another important capability of OCEANS is the generation of standard and ad hoc reports to document and support the design process.

The final major component of the OCEANS software system is its underlying structure. An object-based framework provides the needed flexibility that allows different users to represent, manage, and query the data, construct and use objects, specify and use methods, define output, etc. to best meet their needs. That is, the system should be capable of being easily redefined or reconfigured in order to best represent a specific product, project or company organization, level of expertise, design requirements and considerations, etc. Similarly, the object-based structure must accommodate increasing levels of detail as a design evolves.

In order to effectively support the design and IPPD processes, it is imperative that the system facilitates and encourages communication and collaboration among member of the IPT. It must also easily and seamlessly link to and integrate with methods, models, and tools from other applications.

As mentioned earlier, the most effective way, at this point in the development process, to expand the capability of the OCEANS prototype is through interaction with industry. That is, to apply the prototype system to actual case examples and represent specific problem domains, products, and processes in the system's object-based context. This is best accomplished through collaboration between the developers of OCEANS and industry, including product developers and manufacturers, as well as software developers. The industry interactions not only provide data and expertise on the products being designed and the processes that are used to produce the products, but provide a richer understanding and appreciation of the design and IPPD processes. This understanding, along with feedback from industry as the OCEANS system evolves will result in an application that better meets the commercial and military needs to develop affordable products.

Therefore, a critical next step in the evolution of the OCEANS system is to involve industry partners in order to provide the expertise and feedback that are necessary in order to further refine OCEANS conceptual approaches, populate its object base, and enhance its overall functionality. As illustrated in Figure 15, industry involvement in this endeavor is currently being pursued along two fronts: (1) product/process design and costing and (2) software application design and development. The most likely industry partner for the first front is Northrup-Grumman's Commercial Aircraft Division (NG-CAD). This is based on the author's involvement with the company through the C-17 Manufacturing 2005 Program ([8, 9]). Other potential industry partners that have been identified include The Aerostructures Corporation, an aircraft component manufacturer, and SBIRE, a virtual manufacturing enterprise. Two likely partners from the software industry are, as previously discussed, Cognition Corporation and Knowledge Base Engineering (KBE).

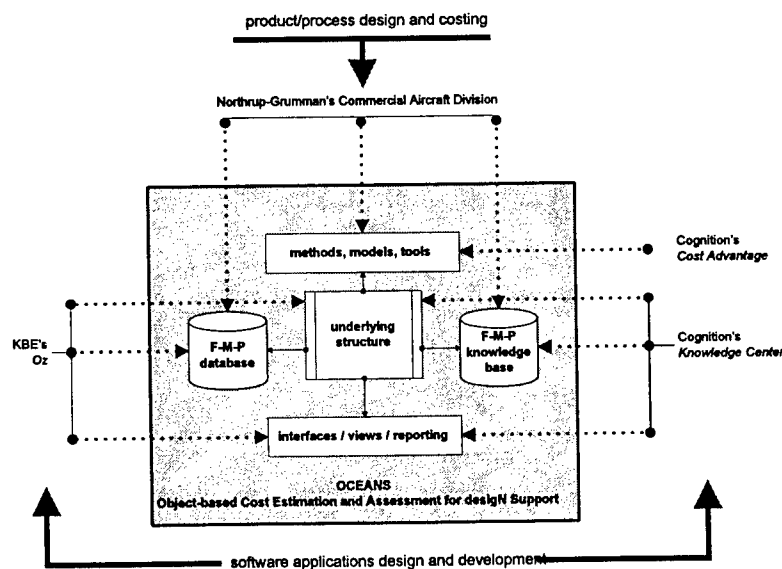


Figure 15: Industrial partners' involvement with future OCEANS systems development.

References

1. Thomas, R., *Manufacturing Cost Estimation, Modeling and Control: A Review of the Literature*, 1994, Stratech, Inc.: Beavercreek, OH.
2. Greenwood, A.G., *A Framework for Manufacturing-Oriented, Design-Directed Cost Estimation*, 1996, Air Force Office of Scientific Research: Bolling Air Force Base, DC.
3. Greenwood, A.G. "An Approach to Enhance Cost Estimation During Product-Process Design," *Proceedings of the Decision Sciences Institute Conference*. 1997. San Diego.
4. Greenwood, A.G., *An Object-Based Approach for Integrating Cost Assessment into Product/Process Design*, 1997, Air Force Office of Scientific Research: Bolling Air Force Base, DC.
5. Khoshafian, S. and R. Abnous, *Object Orientation Concepts, Languages, Databases, User Interfaces*. 1990, New York: John Wiley & Sons, Inc.
6. Noton, B.R., *ICAM Manufacturing Cost/Design Guide (MC/DG) Volume I: Demsonstration Sections*, 1979, Battelle Columbus Laboratories: Columbus, OH.
7. Noton, B.R., *ICAM Manufacturing Cost/Design Guide (MC/DG) Volume II: Appendices to Demsonstration Sections*, 1979, Battelle Columbus Laboratories: Columbus, OH.
8. Fondon, J.W., Jr., et al., *Large Composite Structure -- Commercial Military Integration (Military Products Using Best Commercial Practices) Phase I -- Manufacturing 2005 Business Practices, Manufacturing Infrastructure, Process Technology Development, Volume II -- Manufacturing 2005 Handbook*, . 1996, Manufacturing Technology Directorate, Wright Laboratory, Air Force Materiel Command: Wright-Patterson Air Force Base, OH.
9. Riseley, M.K., T. Houpt, and S. Wheeler, *Large Composite Structure -- Commercial Military Integration (Military Products Using Best Commercial Practices) Phase I -- Manufacturing 2005 Business Practices, Manufacturing Infrastructure, Process Technology Development, Volume I -- Program Review*, . 1996, Manufacturing Technology Directorate, Wright Laboratory, Air Force Materiel Command: Wright-Patterson Air Force Base, OH.

MISMATCH STRESSES, LAMELLAR MICROSTRUCTURE AND MECHANICAL PROPERTIES OF TiAl-ALLOYS

Michael Grinfeld

Research Professor of Mechanical and Aerospace Engineering
Department of Mechanical and Aerospace Engineering

Rutgers University, Piscataway, NJ 08844

Final Report for:
Summer Faculty Research Program
Wright-Patterson Laboratory

Sponsored by:

Air Force Office of Scientific Research
Bolling AFB, DC

December 1997

MISMATCH STRESSES, LAMELLAR MICROSTRUCTURE AND MECHANICAL PROPERTIES OF TiAl-ALLOYS

Michael Grinfeld

Research Professor of Mechanical and Aerospace Engineering
Department of Mechanical and Aerospace Engineering, Rutgers University

Abstract

The elastic misfit stresses in TiAl alloys assume huge level of the order of the product of the elastic moduli times misfit deformations. The elastic energy associated with these stresses is enormously large also and it tries to relax via one of several possible mechanisms. Hence, there appears a driving force for the rearrangement of the material elements leading to the evolution accompanied by the creation of dislocations, migration of vacancies, development of interfacial corrugations, etc...

The goal of this research project is two-fold: i) to develop a theoretical approach which permits a self-consistent calculations of misfit stresses in TiAl lamellar multilayers and like structures, and ii) to develop a thermodynamic approach allowing one to study the possibility of stress relaxation by means of mass rearrangement in the vicinity of the interface between the lamellae. We assume that the rearrangement is accompanied by interfacial diffusion and/or migration of the vacancies. Since the full consideration of the coherent interface solid-solid includes lots of calculation and details we illustrate our results and approach by considering the much simpler but conceptually a very close problem of mass rearrangement in the vicinity of a stress-free interface.

MISMATCH STRESSES, LAMELLAR MICROSTRUCTURE AND MECHANICAL PROPERTIES OF TiAl-ALLOYS

Michael Grinfeld

Contents

1	ORGANIZATION OF THE REPORT	5
2	PUBLICATIONS RESULTING FROM THE GRANT	5
3	INTERNAL STRESSES IN LAMELLAR Ti-Al (physical discussion)	6
3.1	Resume of the Section	6
3.2	Physical motivation and qualitative estimates	6
3.3	Stresses and Strains in Coherent Multilayers in Ti-Al-alloys	7
3.4	Geometry of lamellar Ti-Al	9
4	STRESSES AND STRAINS IN LAMELLAR Ti-Al	10
4.1	Discussion	15
4.2	Conclusions	15
5	STRESSES IN LAMINATED ELASTIC STRUCTURES WITH MISFIT (general thermodynamic study)	16
5.1	The notion and geometry of misfit deformation	16
5.2	Some features of the stresses in laminated structures with misfit	18
5.3	Uniform misfit stresses of the "internal" field	18
5.3.1	General system of equations defining strain-stress field	18
5.4	Misfit stresses in the "free" γ/γ plate-like periodic structure with isotropic moduli	20
5.5	Misfit stresses in the "free" α/γ - periodic structure (different elastic moduli)	21
5.5.1	α/γ - periodic structure (2-layers, isotropic elastic lamellae) . .	21
5.5.2	α/γ - periodic structure (N-layers, anisotropic elastic lamellae) . .	22
5.6	Misfit stresses in the "on-substrate" periodic multilayers with different elastic moduli	23
6	ROLE OF VACANCIES IN SLOW EVOLUTION OF INTERFACES	24
6.1	Introduction	24
6.2	The master system of slow evolution	24

6.3	Rate of change of small disturbances	26
7	CONCLUSION	27
7.1	The main results of this project.	27
7.2	The most promising directions of future research	29

1. ORGANIZATION OF THE REPORT

This report is organized as follows. First, we give an information concerning actual and forthcoming publications and of work in progress supported by this Grant. Then, we present our main results in 3 concentrers which can be read independently although all of them deal with the same object: a solid substance with interfaces. The separation in the 3 concentrers is intended to give access to the problem at hand to materials scientists with different background. First, we present our results relating to misfit strains and stresses in Ti-Al lamellar alloys in the form mostly convenient for materials scientists: both theorists and experimenters. In this section of the report we follow our forthcoming paper "Internal Stresses in Lamellar Ti-Al", prepared in tight collaboration with physicists and materials scientists working at WPL. We skip any lengthy calculations and technical details and concentrate basically on the concepts and formules which seem mostly useful for interested practitioners with background in materials science and mechanical engineering. In order to establish the simplest transparent formules we do not dwell here on any general thermodynamic paradigms, schemes and secondary details. Also, we give preference to the simplest and the most robust models of isotropic lamellae and of misfit deformations. We ignore even the role of temperature and various defects but the most essential lamellar interfaces. All these simplifications allow us to establish quite simple robust formules and qualitative conclusions allowing numerous generalisations. Then, we proceed with presentation of a much more general and fundamental theoretical consideration which was used to establish the basic formules of the previous section and which, hopefully, will become useful for future applications in various topics of materials science and engineering dealing with lamellar microstructure. In the last concentrer we present our results relating to evolution of lamellar microstructure: the presentation is based on the recently published paper "Role of Vacancies in Slow Evolution of a Traction-Free Interface". In order to keep this report reasonably concise, we consider the problem of evolution of lamellar microstructure dealing with the mechanism based on evolution of vacancies. Even this simplest mechanism allows to explain nucleation of dislocations, destruction of alloys, and many other phenomena. Certainly, these phenomena of paramount interest in different branches of materials science and engineering may have several other explanations. We believe though that many of other mechanisms share essential features of the mechanism considered in this report: it is the relaxation of overall stresses based on the development of interfacial corrugations which are known as the Stress Driven Rearrangement Instability.

2. PUBLICATIONS RESULTING FROM THE GRANT

M.A. Grinfeld, P.H. Hazzledine, Role of Vacancies in Slow Evolution of a Traction-Free Interface. *Europhys. Lettr.*, 1997, **37**, pp. 409-413.

M.A. Grinfeld, P.H. Hazzledine, B. Shoykhet, D. M. Dimiduk, Internal Stresses in Lamellar Ti-Al, Metallurgical and Materials Transactions (accepted).

M.A. Grinfeld, Stress Corrosion Cracking in Aggressive Environment, Acta Metallurgica et Materialia (accepted).

B. Shoykhet, M.A. Grinfeld, P.M. Hazzledine, Internal Stresses and Strains in Coherent Multilayers, Acta Metallurgica et Materialia (accepted).

P.M. Hazzledine, B. Shoykhet, M.A. Grinfeld, Coherency Strains and Stresses in Lamellar Ti-Al (Proceedings of Tausier Conference on "Modern Problems of Dislocations").

M.A. Grinfeld, P.H. Hazzledine, B. Shoykhet, Geometrically Nonlinear Theory of Lamellar Microstructure (in preparation).

B. Shoykhet, P.M. Hazzledine, M.A. Grinfeld, Homogenization Methods for Lamellar Microstructure (in preparation).

3. INTERNAL STRESSES IN LAMELLAR Ti-Al (physical discussion)

3.1. Resume of the Section

In this Section we give a coherent physical presentation of the more general results presented in the next Section. We specialize the more general formulae of the next Section for the case of lamellar structure in Ti-Al alloys which were the main concern of our study. General formulae are given for the coherency strains and stresses in a multilayer far from the free surfaces (a detailed derivation of the strain fields, including edge effects, is hard and it is planned to be studied in near future). The multilayer is assumed to be a periodic stack of different elastically isotropic materials but there may be any number of layers in the stack and they may each have any thickness and any elastic constants. The results are applied to lamellar Ti-Al alloys in which there are shear misfits between different γ layers and both shear and biaxial misfits between the γ and α_2 layers. In a fully coherent multilayer the stresses would be large, in the GPa range, and in high strength, thin lamella, alloys the coherency stresses are a substantial fraction of a GPa. The shear stresses act principally on hard mode deformation systems and the biaxial stresses place every α_2 lamella in biaxial compression. This biaxial compression, which, for dislocation glide, is equivalent to a uniaxial tension normal to the lamella is particularly large when the α_2 volume fraction is small.

3.2. Physical motivation and qualitative estimates

When a thin film is deposited epitaxially on a single crystal substrate its lattice parameter initially matches that of the substrate. Consequently the film acquires an elastic strain, which may be large, while the substrate remains almost unstrained. The lattice parameter mismatch which causes the strain generally varies with tem-

perature because expansion coefficients vary from material to material. As the film thickens, the elastic strain remains constant until the energy of the bimaterial can be reduced by the relaxation of the elastic strain through a modification of the interfaces. Substantial elastic strains persist in films which are thicker than this coherence limit [9]. In a multilayer the behavior is similar except that the strains are shared between the different materials. Other things being equal, the thinner layers and those with lower elastic moduli suffer more strain. Most calculations of elastic strains in multilayers have focussed on semiconductor multilayers in which the crystal structures of the different layers are all the same. Consequently the stresses are equi-biaxial tensions and compressions while the strains are triaxial because of the Poisson effect. Multilayers with graded interfaces may be handled by subdividing the multilayer into infinitesimally thick layers [10]. Some multilayers for structural applications are also composed of materials with the same crystal structures and their biaxial tensions and compressions are important in influencing their fracture behavior [1]. In other structural materials, however, the crystal structures do not match and even the symmetries in the layer plane are different; a case in point is Ti-Al. In lamellar Ti-Al alloys a hexagonal α_2 phase and a tetragonal γ phase are joined on their close packed planes. The α_2 lattice parameter is larger than the lattice repeat of γ so biaxial tensions and compressions, similar to those in semiconductors are necessary. In addition, the α_2 structure has six-fold symmetry while the γ structure lacks even three-fold symmetry so shears are necessary to achieve coherency. The Ti-Al alloy multilayer has a wide spread of layer thicknesses. In order to calculate the coherency stresses a theory is required which can handle large numbers of layers, of widely differing thicknesses, different sets of elastic constants and different lattice repeats. A first estimate of the stresses [13] showed that in polysynthetically twinned (PST) material with 1mm thick layers, stresses of the order of 50MPa would be expected and these have been detected experimentally [13], [2]. In PST the coherency stresses are a small fraction of the yield stress. The motive for this much more elaborate calculation is that the new high strength Ti-Al alloys have much finer lamellae and consequently much larger coherency stresses. In a partially coherent multilayer like Ti-Al, with layer thickness H , the coherency stresses fall as H^{-1} while the yield stress falls as $H^{-1/2}$ so the coherency stresses become increasingly significant as the layers thin. When H is about 100nm, the calculations in this paper show that the largest coherency stress is as large as the yield stress and is therefore capable of markedly influencing the mechanical properties.

3.3. Stresses and Strains in Coherent Multilayers in Ti-Al-alloys

Consider a multilayer consisting of a stack of flat layers of different materials, with uniform thicknesses and planar dimensions very much larger than their thicknesses (see, Figure 1). Each layer may have different elastic constants, assumed to be isotropic. The layers, which are parallel to the x^3 plane, have individual thicknesses H^k and are arranged in a periodic stack with N members. N may take any value and the total

thickness of the periodic stack is H . The fact that the individual layers are different materials implies that their lattice constants would not match if they were not joined. When they are joined in a fully coherent multilayer, all the lattice parameters in the plane of the layers (but not perpendicular to it) must be the same, consequently elastic strains and stresses exist in every layer. The fact that there is no requirement for matching along x^3 (the layers can expand or contract along x^3 freely) indicates that this is a plane stress problem, $P^{33} = 0$. Clearly the magnitude of the elastic strains depends on the initial misfit, Δ , defined as the strain required to transform some reference cell into the cell of the particular free layer. In the coherent state, each layer may be thought to contain two strains, its Δ and its elastic strain K . The sum of these two ($\Delta + K$) must be the same in every layer at coherency. The aim is to calculate the elastic strain in each of the N layers in the period of the multilayer. In principle, N may be very large but a physical argument shows that it is the number of crystallographic variants which determines the size of the calculation, rather than N . In a Ti-Al multilayer containing N layers in a period there are only four distinct variants (a variant has a unique combination of Δ , and n where m is the shear modulus and n the Poisson ratio). In the coherent multilayer, all the lattice parameters match so each layer belonging to a given variant must have the same strain K and therefore the same stress P . It makes no difference to the argument where the individual members of a variant are located in the multilayer so they may, for convenience, be collected into a single 'ply'. The problem of calculating the strains and stresses in a Ti-Al multilayer containing thousands of layers is then reduced to the problem of calculating the strains and stresses in a four-ply multilayer. It should be noted, however, that this is a conceptual process rather than a real one because individual layers must be very thin to be coherent. By the time the layers have been collected into a ply, its thickness may be far above the limiting thickness for coherency so a 'conceptual' four-ply has larger coherency strains in it than a real four-ply. At the boundaries of the multilayer, the strains vary over a distance comparable with the thickness of a layer but far from the edges they are uniform. In this paper we examine the uniform strains within each layer of the multilayer, strains which necessarily have a discontinuity at every interface to ensure coherency. The fields obey the normal rules of strain compatibility and stress equilibrium within a layer and they provide mechanical equilibrium at each interface. The strains scale in such a way that the volume average of each stress component is zero, as it must be in a solid with no external loads. The zero volume average stress condition brings the multilayer to a state of minimum strain energy.

In order to present the results of the next Section in the most transparent form the following notions and notation are required. We denote N the total number of the layers in the stack with individual thickness H_k , shear modulus μ_k , Poisson ratio ν_k , and biaxial modulus $M_k = 2\mu_k(1 + \nu_k)/(1 - \nu_k)$. Two additional average parameters of the stack are required. The first is an integrated shear modulus $S = \sum H_k \mu_k$ and the second is an integrated biaxial modulus $B = \sum H_k M_k$. The symmetric strain tensor $K_L^{(ij)}$ in layer L is then given as a function of material parameters, S , B and

transformation strains Δ as follows:

$$K_L^{(ij)} = \frac{1}{2S} \sum_{M=L}^N H_M \mu_M (2\delta_{i\pi} \delta_{j\omega} - \delta_{ij} \delta_{\pi\omega} + \delta_{\pi\omega} n_i n_j) (\Delta_M^{(\pi\omega)} - \Delta_L^{(\pi\omega)}) \quad (3.1)$$

$$\frac{1}{2B} \sum_{M=L}^N H_M M_M \left(\delta_{ij} \delta_{\pi\omega} - \frac{1+\nu_L}{1-\nu_L} \delta_{\pi\omega} n_i n_j \right) (\Delta_M^{(\pi\omega)} - \Delta_L^{(\pi\omega)})$$

In terms of the same parameters, the stress tensor $P_L^{\alpha\beta}$ in layer L is

$$P_L^{\alpha\beta} = \left\{ \frac{\mu_L}{2S} \sum_{M=L}^N H_M \mu_M (2\delta_\pi^\alpha \delta_\omega^\beta - \delta^{\alpha\beta} \delta_{\pi\omega}) + \frac{M_L}{2B} \sum_{M=L}^N H_M M_M \delta^{\alpha\beta} \delta_{\pi\omega} \right\} (\Delta_M^{(\pi\omega)} - \Delta_L^{(\pi\omega)}) \quad (3.2)$$

In both eqs. 3.1 and 3.2, roman indices run from 1 to 3 while greek indices run from 1 to 2. $n_i = \delta_i^3$ and δ is a Kronecker delta. The meaning of $\Delta_M^{(\pi\omega)} - \Delta_L^{(\pi\omega)}$ is the strain required to transform the cell of layer L to the cell of layer M in the free state.

3.4. Geometry of lamellar Ti-Al

The form of the multilayers in Ti-Al is shown in Figure 2. The lamellar structure is derived from a grain of α_2 by an incomplete transformation from the hexagonal DO_{19} phase to the tetragonal γ phase with the $L1_0$ structure. The α_2 lamellae are the remnants of the original α_2 grain and consequently they all have the same crystallographic orientation. The plane of the α_2 lamellae is the (0001) basal plane and the plane of the γ lamellae is {111}. Close packed directions in all the crystals are approximately aligned, consequently there are six rotational variants of the γ lamellae (labelled $\gamma_0, \gamma_{60}, \dots, \gamma_{300}$, the rotation is anticlockwise with respect to variant 0) of which four are shown in figure 2. Lamellar Ti-Al alloys take two forms: Either they are derived from a single α_2 grain and referred to as polysynthetically twinned (PST) or they are derived from polycrystalline α_2 and referred to as fully lamellar (FL). In both cases, within one lamellar colony, a small fraction of the lamellae are continuous α_2 crystals and the remainder are γ crystals which are divided into domains (themselves different rotational variants of γ), see Figure 2. Typically, the grain size of FL material or the dimensions of a PST specimen are large, of order mm or cm. The domain size is typically 50nm and the lamellar thicknesses are in the range 20 nm to 1mm. In all cases, therefore, the crystals have much larger dimensions in the plane of the lamellae than normal to it. The $L1_0$ structure is most conveniently described as face centred cubic with alternate {002} planes occupied by Ti and Al atoms. If the c/a ratio of this unit cell were 1.0, all of the γ variants would be fully coherent with one another but they would not be coherent with α_2 because the lattice parameter a of α_2 is larger than $a\sqrt{2}$ in γ . Coherency could be achieved by the presence of biaxial tensions in γ and biaxial compressions in α_2 . In fact, c/a of γ is about 1.02 and this

implies that only twinned pairs of γ variants match on the $\{111\}$ planes (γ_0 matches with γ_{180} , γ_{60} matches with γ_{240} etc.). Shear strains are required in the γ lamellae to make them fully coherent. In a γ/α_2 material, both biaxial tension/compression and shears are needed. In the theory of section 2, the coherency strains contain Δ , the strain required to transform the dimensions of a unit cell of a reference lattice into the dimensions of the unit cell of a free lamella. In the theory only the in-plane dimensions are matched, the unit cells may differ in length along the x^3 direction. In principle, any unit cell could be used as the reference lattice (because all the formulae contain differences between two Δ tensors) but here a choice is made to use the unit cell of the γ_0 variant (or, equally, the γ_{180}) as the reference, so $\Delta_0 = \Delta_{180} = 0$. The remaining transformation strains are defined as the strains required to transform the γ_0 cell into the cell in question. Setting x^1 parallel to $\langle 110 \rangle$ in the γ_0 variant, the transformation strains are:

$$\Delta_{60} = \Delta_{240} = \frac{\varepsilon}{2} \begin{bmatrix} 1 & -1/\sqrt{3} \\ -1/\sqrt{3} & -1 \end{bmatrix} \quad (3.3)$$

$$\Delta_{120} = \Delta_{300} = \frac{\varepsilon}{2} \begin{bmatrix} 1 & 1/\sqrt{3} \\ 1/\sqrt{3} & -1 \end{bmatrix} \quad (3.4)$$

$$\Delta_{\alpha 2} = \begin{bmatrix} \varepsilon' & 0 \\ 0 & \varepsilon' - 2\varepsilon/3 \end{bmatrix} \quad (3.5)$$

where $\varepsilon = (c/a - 1)$ in the γ phase and $\varepsilon' = (a_{\alpha_2} - \sqrt{2}a_\gamma)/a_{\alpha_2}$ is one of the mismatches between α_2 and γ .

4. STRESSES AND STRAINS IN LAMELLAR Ti-Al

The stresses and strains in any layer of a coherent Ti-Al multilayer can be found from equations 3.1 and 3.2 so long as the volume fractions of the four components, the elastic constants and the lattice parameters are known. In this section, the stresses and strains in three idealised forms of Ti-Al are given explicitly to act as a guide for the discussion in section 5. The idealised forms are (i) Ti-Al which contains only the three gamma variants, (ii) Ti-Al which contains alternating layers of α_2 and a single γ variant, γ_0 and (iii) Ti-Al which contains γ_0 , γ_{60} and γ_{120} in equal volume fractions f and α_2 at volume fraction $(1 - 3f)$. The first is a good approximation to many PST specimens which contain α_2 concentrations of about 10% in high strength alloys and the third is a realistic representation of lamellar Ti-Al since there is no obvious reason why one variant of γ should be more common than another.

All γ alloys

The material may contain all six variants of γ but the strains are the same in twinned pairs so only three strain fields exist. Since isotropic elasticity is assumed, the elastic constants are the same for each lamella. This is a plane stress problem, so $p_{33} = 0$ but inspection of the transformation strains, eq. ?? and ??, shows that only in-plane shears exist therefore $K_{33} = 0$ also. Substitution of the transformation

strains, ?? and ?? into 3.1 give the following strains, K_0 in the 0° and 180° variant layers, K_{60} in the 60° and 240° variant layers and K_{120} in the 120° and 300° variant layers:

$$\frac{2}{\epsilon}K_0 = f_{60} \begin{bmatrix} 1 & -1/\sqrt{3} \\ -1/\sqrt{3} & -1 \end{bmatrix} + f_{120} \begin{bmatrix} 1 & 1/\sqrt{3} \\ 1/\sqrt{3} & -1 \end{bmatrix} \rightarrow \begin{bmatrix} 2/3 & 0 \\ 0 & -2/3 \end{bmatrix} \quad (4.1)$$

$$\frac{2}{\epsilon}K_{60} = f_0 \begin{bmatrix} -1 & 1/\sqrt{3} \\ 1/\sqrt{3} & 1 \end{bmatrix} + f_{120} \begin{bmatrix} 0 & 2/\sqrt{3} \\ 2/\sqrt{3} & 0 \end{bmatrix} \rightarrow \begin{bmatrix} -1/3 & 1/\sqrt{3} \\ 1/\sqrt{3} & 1/3 \end{bmatrix} \quad (4.2)$$

$$\frac{2}{\epsilon}K_{120} = f_0 \begin{bmatrix} -1 & -1/\sqrt{3} \\ -1/\sqrt{3} & 1 \end{bmatrix} + f_{60} \begin{bmatrix} 0 & -2/\sqrt{3} \\ -2/\sqrt{3} & 0 \end{bmatrix} \rightarrow \begin{bmatrix} -1/3 & -1/\sqrt{3} \\ -1/\sqrt{3} & 1/3 \end{bmatrix} \quad (4.3)$$

The strains on the right side are the limiting values when all six variants have equal volume fractions, $f_0 = f_{60} = f_{120} = 1/3$.

Since $K_{33} = 0$ in all cases, the stress fields have identical forms:

$$P_0 = 2\mu K_0 \quad P_{60} = 2\mu K_{60} \quad P_{120} = 2\mu K_{120} \quad (4.4)$$

In the coherent multilayer, the in-plane lattice parameters all match each other. Referring to Fig. 2, the triangles have sides parallel to close packed directions and the lengths of these sides illustrate the effects of the coherency strains: In the γ_0 variant, viewed down $(-1 - 1 - 1)$ the unique side, $[-110]$, has a free length $a\sqrt{2}$ and the two other sides have lengths $a\sqrt{2}(1 + \epsilon/2)$. In the coherent state, the $[-110]$ side assumes a length of $a\sqrt{2}[1 + (f_{60} + f_{120})\epsilon/2]$, the $[-101]$ side has length $a\sqrt{2}[1 + (f_0 + f_{120})\epsilon/2]$ and the $[0-11]$ side has length $a\sqrt{2}[1 + (f_0 + f_{60})\epsilon/2]$. If $f_0 = f_{60} = f_{120} = 1/3$, all three sides have the same length $a\sqrt{2}(1 + \epsilon/3)$. Formally the γ phase loses most of its symmetry in the coherent multilayer and becomes triclinic, only regaining trigonal symmetry when the volume fractions of the variants are equal. In the 'all gamma' approximation to lamellar Ti-Al, the coherency strains and stresses are entirely homogeneous shears in the plane of the lamellae. The typical magnitude of the shear stress is $2\mu\epsilon/3$ or about 0.9 GPa in a fully coherent material with $\mu = 70$ GPa and $\epsilon = 0.02$. The orientation of the stress is such that it exerts no force on soft mode dislocations and can therefore not influence the plastic behavior of the soft mode. Hard mode deformation systems may be either softened or hardened by the coherency stresses.

α_2 - single γ variant bilayered material

The material under consideration here consists of alternate layers of the γ_0 variant ($\Delta_1 = 0$) and α_2 . The transformation strain for α_2 given in eq. ?? only has principal strains Δ^{11} and Δ^{22} because x_1 and x_2 are principal axes. If rotated axes had been used there would be a shear strain Δ^{12} and this is included for completeness. The transformation strain used is:

$$\Delta_2 = \begin{bmatrix} \Delta^{11} & \Delta^{12} \\ \Delta^{21} & \Delta^{22} \end{bmatrix} \quad (4.5)$$

Subscript 1 refers to material 1, that is γ_0 , and subscript 2 refers to material 2, that is α_2 . Writing M for the biaxial modulus $2\mu(1+\nu)/(1-\nu)$, the strains in the two materials are:

$$K_1^{11} = \frac{f_2}{2} [M_2 (\Delta^{11} + \Delta^{22}) / M^* + \mu_2 (\Delta^{11} - \Delta^{22}) / \mu^*] \quad (4.6)$$

$$\frac{f}{2} [M_2 (\varepsilon' - \varepsilon/3) / M^* + \mu_2 \varepsilon / 3\mu^*]$$

$$K_1^{12} = K_1^{21} = \frac{f_2 \mu_2 \Delta^{12}}{2\mu^*} \quad (4.7)$$

$$K_1^{22} = \frac{f_2}{2} [M_2 (\Delta^{11} + \Delta^{22}) / M^* - \mu_2 (\Delta^{11} - \Delta^{22}) / \mu^*] \quad (4.8)$$

$$f [M_2 (\varepsilon' - \varepsilon/3) / M^* - \mu_2 \varepsilon / 3\mu^*]$$

$$K_1^{33} = -f_2 \nu_1 X (\Delta^{11} + \Delta^{22}) / 2\mu_1 (1 + \nu_1) \quad (4.9)$$

$$-f_2 \nu_1 X (\varepsilon' - \varepsilon/3) / \mu_1 (1 + \nu_1)$$

where M^* is the average biaxial modulus $f_1 M_1 + f_2 M_2$, μ^* is the average shear modulus $f_1 \mu_1 + f_2 \mu_2$ and where $X = M_1 M_2 / M^*$. The first forms in eq. 4.6-4.9 are the general forms and the second are the particular forms for γ_0/α_2 i.e. when $\Delta^{11} = \varepsilon'$, $\Delta^{12} = \Delta^{21} = 0$ and $\Delta^{22} = (\varepsilon' - 2\varepsilon/3)$.

The strains in layer 2 are obtained by interchanging the subscripts on the material parameters f , μ , ν , M and multiplying by -1. The stresses in the γ_0 layer of γ_0/α_2 are:

$$P_1^{11} = f_2 \left[X \frac{\Delta^{11} + \Delta^{22}}{2} + Y \frac{\Delta^{11} - \Delta^{22}}{2} \right] \rightarrow f_2 [X (\varepsilon' - \varepsilon/3) + Y \varepsilon/3] \quad (4.10)$$

$$P_1^{12} = P_1^{21} = f_2 Y \Delta^{12} \quad (4.11)$$

$$P_1^{22} = f_2 \left[X \frac{\Delta^{11} + \Delta^{22}}{2} - Y \frac{\Delta^{11} - \Delta^{22}}{2} \right] \rightarrow f_2 [X (\varepsilon' - \varepsilon/3) - Y \varepsilon/3] \quad (4.12)$$

$$P_1^{33} = 0 \quad (4.13)$$

where $Y = 2\mu_1 \mu_2 / \mu^*$

Again, the first form is the general form and the second form is the particular result for γ_0/α_2 . The stresses in layer 2 are $P_2^{\alpha\beta} = -(f_1/f_2)P_1^{\alpha\beta}$ as they must be to comply with the zero volume average of stress components rule. In the γ_0/α_2 approximation to lamellar Ti-Al there are three types of stresses with different origins (i) in-plane shear stresses, caused by the tetragonality of γ , (ii) out-of-plane shear stresses and

(iii) biaxial tension/compression stresses caused by the fact that α_2 has larger lattice parameters than γ .

The in-plane shear stresses have the same origin (the tetragonality of γ) as above and consequently the same order of magnitude. But this stress is split between the phases and in a typical Ti-Al alloy in which the α_2 volume fraction is 0.1, the α_2 layers have a stress of about 1.6 GPa and the γ layers a stress of about 0.2 GPa. However, this conclusion is reached because the γ layers are all of one variant. If there were a mixture of variants, the in-plane shear stresses would be lower in α_2 and higher in γ (see below).

The out-of-plane shear stresses have maximum values of $P_{11}/2$ and $P_{22}/2$. The larger of these is $P_{11}/2$ in both layers because there is a larger misfit between α_2 and γ_0 in the x^1 direction than there is in the x^2 direction. The value of this shear stress in α_2 is about 2.3 GPa and in γ_0 it is about 0.3 GPa. The reason for the disparity is, as before, that the minority phase strains more than the majority phase (when the elastic constants are comparable) and consequently is more highly stressed. The biaxial compression in the α_2 layers (and consequent biaxial tension in the γ layers) deserves a detailed discussion because it could have a decisive effect on the mechanical properties of Ti-Al alloys. The fact that $P_2^{11} + P_2^{22}$ is negative in α_2 layers gives this material a substantial protection against the propagation of a crack across the lamella which it would not have in the monolithic form. The larger the lattice parameter of α_2 can be made by alloying, the greater this protection will become. The biaxial compressive stress in α_2 exerts large forces on dislocations gliding on the pyramidal (but not the basal or prismatic) slip planes. Since dislocation glide is unaffected by hydrostatic pressure, it is easier to think of the biaxial compression as an effective uniaxial tension along the basal plane normal: the average compression is $(P_2^{11} + P_2^{22})/2$. If a pressure of this magnitude is subtracted from each of the normal stresses in α_2 the effective stresses Π are

$$\Pi^{11} = \frac{P_2^{11} - P_2^{22}}{2} \quad \Pi^{22} = \frac{P_2^{22} - P_2^{11}}{2} \quad \Pi^{33} = -\frac{P_2^{11} + P_2^{22}}{2}$$

Π^{33} is the effective tensile stress along [001]. Its magnitude is approximately $4.2f_1$ GPa. There is a compensating compressive stress of $-4.2f_2$ GPa in the γ layers. When $f_1 \gg f_2$ the very large effective tensile stress in α_2 must make it much more easily deformable in tension along [0001] than in compression and PST specimens deformed in the N orientation would be expected to show a substantial tension/compression asymmetry if slip on the pyramidal planes of α_2 is important in the deformation.

α_2 - three γ variant alloys

In this example the three (or six) variants of γ are equally represented with total volume fraction $3f$, the remaining $(1 - 3f)$ being α_2 . Layer 1 is γ_0 and γ_{180} , layer 2 is γ_{60} and γ_{240} , layer 3 is γ_{120} and γ_{300} and layer 4 is α_2 .

The strains are:

$$K_1 = K + \frac{\varepsilon}{2} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad (4.14)$$

$$K_2 = K + \frac{\varepsilon}{3} \begin{bmatrix} -1/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix} \quad (4.15)$$

$$K_1 = K + \frac{\varepsilon}{2} \begin{bmatrix} -1/2 & -\sqrt{3}/2 \\ -\sqrt{3}/2 & 1/2 \end{bmatrix} \quad (4.16)$$

where

$$K = [M_2(1-3f)(\varepsilon' - \varepsilon/3) M^*] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\frac{2\nu_1}{1-\nu_1} \end{bmatrix} \quad (4.17)$$

$$K_4 = [M_1(1-3f)(\varepsilon' - \varepsilon/3) M^*] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\frac{2\nu_2}{1-\nu_2} \end{bmatrix}$$

The stresses are

$$P_1 = X(1-3f)(\varepsilon' - \varepsilon/3) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \mu_1 \varepsilon \begin{bmatrix} 2/3 & 0 \\ 0 & -2/3 \end{bmatrix} \quad (4.18)$$

$$P_2 = X(1-3f)(\varepsilon' - \varepsilon/3) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \mu_1 \varepsilon \begin{bmatrix} -1/3 & 1/\sqrt{3} \\ 1/\sqrt{3} & 1/3 \end{bmatrix} \quad (4.19)$$

$$P_3 = X(1-3f)(\varepsilon' - \varepsilon/3) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \mu_1 \varepsilon \begin{bmatrix} -1/3 & -1/\sqrt{3} \\ -1/\sqrt{3} & 1/3 \end{bmatrix} \quad (4.20)$$

$$P_1 = -X(1-3f)(\varepsilon' - \varepsilon/3) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (4.21)$$

These strains and stresses have some of the features of both the previous examples. The γ variants may be thought to reach coherency in two stages: First they shear by the same amount as in eq. ??, reaching three-fold symmetry because the volume fractions of the three variants are equal; second, they undergo a biaxial stretch represented by K . The α_2 layers suffer a biaxial compression which is large when f is close to $1/3$ i.e. when α_2 is a minority phase. At coherency, the lattice parameters in the x^3 plane are equal, with value $a\sqrt{2} = (1 + K_1^{11})$:

$$a_c = a\sqrt{2}[1 + 3fM_1\frac{\varepsilon}{3} + (1-3f)M_2(\varepsilon')]/M^* \quad (4.22)$$

which is the weighted mean of the individual lattice parameters, the weighting factor being M . Explicitly,

$$a_c = \frac{a_2}{M^*} [fM_1 + fM_1(1 + \varepsilon/2) + fM_1(1 + \varepsilon/2) + (1-3f)M_2(1 + \varepsilon')]$$

which is the same as 4.22. The result in 4.22 is the most compact form of the elastic solution to the entire multilayer since, from it, the strains of 4.14-4.17 may be obtained directly and the stresses of 4.18-4.21 may be obtained from ?? by Hookes law.

4.1. Discussion

The stresses and strains in a fully coherent γ/α_2 lamellar Ti-Al alloy depend on the crystallographic mismatches ϵ , elastic constants and volume fractions of the γ variants and α_2 . Experimentally, the three variants of γ are usually equally common and the volume fraction of α_2 lies between 0.1 and 0.5. Both ϵ and ϵ' depend on composition. The γ phase is close to stoichiometry [26], [7] and ϵ is measured to be close to 0.02 at room temperature [8]. ϵ' is less accurately known since the composition of α_2 is far from stoichiometric Ti₃Al. In our estimates we have assumed it to be 0.021, as given by one set of measurements [23]. Changes of temperature can change ϵ and ϵ' because expansion coefficients in the basal plane of α_2 and along x and z in γ are all different.

Taking ϵ to be 0.02 and ϵ' to be 0.021, the most significant stresses are in-plane shears (proportional to ϵ) in all the γ lamellae of about 0.9 GPa and a biaxial compression (proportional to $\epsilon - \epsilon'/3$) of between 3.8 GPa and 2.1 GPa in α_2 and a compensating biaxial tension in γ of between 0.4 GPa and 2.1 GPa. The biaxial stresses vary between these limits as the volume fraction of α_2 varies from 0.1 to 0.5.

These estimates of the internal stresses are upper limits because Ti-Al has lamellae which are not thin enough to be fully coherent. Much of the stress is relaxed by interfacial dislocations and the measured stresses are one to two orders of magnitude smaller than the coherent estimates above [13], [2]. To our knowledge, there is no detailed theory to determine the coherency limit in a multilayer, especially one containing lamellae with a variety of thicknesses. But the Frank and van der Merwe theory [9] can be used to make an estimate of the maximum lamellar thickness H_c in a coherent multilayer and also the factor by which the stresses are reduced in partially coherent multilayers. The theory shows [13] that in Ti-Al, H_c is approximately 30 nm and in lamellae which are thicker, the stresses are reduced by a factor H_c/H . In PST material, the lamellae are typically 1 μ m thick and the stresses are consequently reduced from the coherent values by a factor of about 30, in line with experiment. In recently developed high strength alloys the lamellae are typically 100 nm thick and, in them, the stresses are predicted to be reduced from the coherent values by only a factor of 3. The coherency stresses in high-strength lamellar Ti-Al alloys are therefore certainly sufficient to contribute significantly to their yield stress of approximately 1 GPa.

4.2. Conclusions

1. General formulae are given for the stress and strain tensors in any layer of a coherent multilayer which may contain any number of different layers, with different thicknesses and different elastic constants. The restrictions in the calculation are that isotropic elasticity is assumed, only in-plane mismatches are used and free surface effects, though taken into account, are not included in the formulae.

2. At coherency all the lattice parameters match so each layer of a particular material or variant has undergone the same strain and contains the same stress. In an

ideal Ti-Al multilayer there are only four coherency stress tensors and four coherency strain tensors in the whole multilayer.

3. In a bimaterial multilayer, each stress component in one material is proportional to the volume fraction of the second material. In Ti-Al the minor phase is α_2 and consequently the stresses and strains are higher in this phase than in γ .

4. In a Ti-Al multilayer containing only g variants, the in-plane lattice parameters are the averages of the individual γ lattice parameters. When the volume fractions of the variants are equal, the coherent multilayer shows three fold symmetry on the layer plane.

5. The in-plane shear stresses in Ti-Al, caused by the tetragonality of the unit cell, do not influence the soft mode of deformation but do exert forces on hard mode dislocations.

6. In γ/α_2 coherent multilayers the lattice parameter is the weighted mean of the individual lattice parameters, the weighting factor being M , the biaxial modulus.

7. In γ/α_2 multilayers large biaxial compressive stresses exist in the α_2 layers. These stresses tend to prevent crack propagation through α_2 and they promote the tensile plastic deformability of α_2 normal to the lamellar plane.

5. STRESSES IN LAMINATED ELASTIC STRUCTURES WITH MISFIT (general thermodynamic study)

5.1. The notion and geometry of misfit deformation

In this Section we give a simplified linear description of elastic heterogeneous systems with coherent interfaces and misfit strains (see Dimiduk et al [6], [7], Hazzledine et al [13], Kad et al [15] for the crystallographic description and physical motivation, and Grinfeld [11] for the more general situation and a more detailed description of the mechanics of heterogeneous solids with finite misfit deformations). Let us choose one of the unstressed configurations of the crystalline substance to be the reference basis for the lamellae elastic structure. In the reference configuration we choose some affine coordinate system x^i ; the Latin indices run the values 1, 2, 3; the summation convention over repeated indexes is implied in what follows. We remind the reader that each affine coordinate system has straight coordinate axes which are not necessarily perpendicular to each other (in particular, any Cartesian coordinate system is necessarily affine): in the following we specify the choice of a specific coordinate system which is the most convenient for the structure in question.

We denote by n^i the components of a unit vector which is perpendicular to the planes along which the lamellae match each other. The total displacement within L -th each of the lamellae $U_i(x^k)$ can be splitted into two parts

$$U_i(x^k) = \Delta_{ij} x^j + u_i(x^k) \quad (5.1)$$

where the first term $\Delta_{ij} x^j$ is associated with the misfit deformations whereas the second $u_i(x^k)$ is associated with the additional elastic deformations. We assume that the misfit strain Δ_{ij} itself does not generate any stresses in the sense that under this strain the isolated lamellae does not accumulate any elastic energy. The misfit strain, however, is the indirect source of huge elastic stresses within the laminated structure since the strains Δ_{ij} differ very significantly (as compared with conventional elastic deformations) within different lamellae. That difference in the misfit strain for different lamellae triggers paramount elastic strains $\partial_j u_i(x^k)$ and stresses p^{ij} because of the requirement of coherency across the interfaces between different lamellae. In what follows the coherency is implied as the continuity of the total displacement across the interfaces. Using the notation $[A]_K^{K+1}$ of the jump of the quantity A across the interface separating the lamellae numbered as $K+1$ and K we can present the coherency condition as follows:

$$[U_i]_K^{K+1} = [\Delta_{ij} x^j + u_i(x^k)]_K^{K+1} = 0 \quad (5.2)$$

If we choose the coordinate system in such a way that the coordinate planes $x^1 - x^2$ (i.e., the plane $x^3 = \text{const}$) coincide with the interfaces between the lamellae, the equation 5.2 appears to be equivalent to the following one:

$$[\Delta_{ij} + \partial_j u_i(x^k)]_K^{K+1} \delta_\alpha^i = 0 \quad (5.3)$$

Here and in what follows the Greek indices run the values 1, 2 and the summation convention over repeated Greek indices is implied; the "shift" symbol δ_α^i is equal to 1 if $i = \alpha$ and it is equal to zero otherwise.

The equality of the forces across the coherent interface can be written as

$$[p^{ij}]_K^{K+1} n_j = 0 \quad (5.4)$$

The assumption of the linear elasticity gives the following formula:

$$p_L^{ij} = c_L^{ijkl} \partial_l u_k \quad (5.5)$$

where the elasticity tensors are symmetric with respect to the permutation of the indices $c_L^{ijkl} = c_L^{jikl} = c_L^{klij} = \dots$. Combining 5.4, 5.5, we arrive at the following equations:

$$[c^{ijkl} \partial_l u_k]_K^{K+1} n_j = 0 \quad (5.6)$$

Within the bulk of each lamellae the following equilibrium equation should be satisfied:

$$c_L^{ijkl} \partial_k \partial_l u_k = 0 \quad (5.7)$$

The system of equations 5.3, 5.6, and 5.7 allows to find equilibrium displacement within each of the lamellae $u_L(x^k)$ provided the appropriate mechanical conditions at the external interface is specified.

The equilibrium displacements depend significantly on the "misfit" strains Δ_{ij} and the elasticity tensors c_L^{ijkl} . The following two experimental observations allow to simplify the system of equilibrium equation significantly:

- i) the elasticities c_L^{ijkl} are the same within each of the lamellae; in the following consideration we use for them the notation c^{ijkl} ,
- ii) only the in-plane components $\Delta_{\alpha\beta}$ of the misfit strain Δ_{ij} deviate from zero; hence

$$\Delta_{ij} = \Delta_{\alpha\beta} \delta_i^\alpha \delta_j^\beta, \quad \Delta_{i\alpha} = 0 \quad (5.8)$$

5.2. Some features of the stresses in laminated structures with misfit

The TiAl alloys often have a laminated structure with the laminae having the in-plane dimensions much greater as compared with their thickness. Because of this very feature some general conclusions regarding the equilibrium stress-strain fields can be made in advance (these conclusions can be made on the basis of both: the experiment and relevant asymptotic analysis of the elasticity equations). The equilibrium fields can be splitted into two parts: i) the uniform stress-strain fields on certain distance from the external boundary and the external edges comparable with the thicknesses of individual lamellae (in the following we name it the "internal" field), and ii) the edge distortions of the above "internal" field which appear to be very significant in the vicinity of the external interfaces (in the following we name it the "external" field). The conditions at the external interface cannot change the very fact of the uniform character of the "internal" field though they can change the specific parameters of the "internal" field; needless to say, that the conditions at the external surface play a crucial role in the formation of the "external" stress-strain field. Both - the "external" and the "internal" fields can play significant roles in different aspects of the behavior of the lamellae-like alloys.

5.3. Uniform misfit stresses of the "internal" field

5.3.1. General system of equations defining strain-stress field

The equilibrium "internal" uniform displacements within the plates $u_L(x^k)$ can be written as follows:

$$u_L(x^k) = \kappa_{ij} x^j \quad (5.9)$$

where is the unknown matrix of the equilibrium gradients. Inserting 5.9 in the equations 5.3 and 5.6 we get, respectively:

$$[\kappa_{ij}]_K^{K+1} \delta_\alpha^i = -[\Delta_{ij}]_K^{K+1} \delta_\alpha^i \quad (5.10)$$

$$[c^{ijkl} \kappa_{kl}]_K^{K+1} n_j = 0 \quad (5.11)$$

The system of algebraic equations 5.10-5.11 includes less equations, than the amount of the unknowns κ_{ij} . This system includes only the demands based on the mechanical equilibrium and coherency at the interfaces. To make the system self-sufficient certain additional physical ideas should be implemented. A deeper analysis shows that there are several "natural" extensions of the system 5.10-5.11. Here we shall discuss only two extensions which are of interest for TiAl multilayers. One of them concerns the "misfit" stresses in multilayers grown on a thick substrate, the other - the "misfit" stresses in the multilayer which are free from any external support. In the former case the substrate appears to be unstressed because of its thickness: we refer the misfit stresses in this case as the "on-substrate" misfit stresses. We counterpose these "on-substrate" misfit stresses to the "free" misfit stresses related to the latter case.

The case of the "free" misfit stresses is more difficult for the calculations and it will be studied first. For this very case and periodically repeated stocks of the lamellae the appropriate asymptotic analysis shows that for the following equation should be satisfied [24]:

$$\sum_{K=1}^N \frac{H}{K} p^{ij} = \sum_{K=1}^N \frac{H}{K} c^{ijkl} \kappa_{kl} = 0 \quad (5.12)$$

where $\frac{H}{K}$ is the thickness the K -th lamella within the period.

The $6(N-1)$ equations 5.10, $3(N-1)$ equations 5.11, and 9 5.12 form a closed system of equations with respect to $9N$ unknowns κ_{kl} . By using the equation 5.10 we can find $3N$ unknowns corresponding to the skew-symmetric (rotational) part of the gradients $\kappa_{[kl]}$. Calculation gives the following formulae:

$$[\kappa_{[ij]}]_K^{K+1} = [\Delta_{[\alpha\beta]}]_K^{K+1} \delta_i^\alpha \delta_j^\beta + 2 [\kappa_{(k\alpha)}]_K^{K+1} n^k \delta_i^\alpha n_j \quad (5.13)$$

The equation 5.13 allows to calculate the equilibrium rotations $\kappa_{[ij]}$ within each individual lamellae provided the equilibrium deformations $\kappa_{(k\alpha)}$ have already been determined from some other equations exhibited below. The equation 5.13 shows that i) the equilibrium rotation within the laminae are actually generated by the rotations in the misfit deformations and by the rotations associated with the deformations, and ii) the equilibrium displacements can be determined to within an arbitrary rotation of the sample as the whole.

For the $6N$ unknowns corresponding to the symmetric (deformation) part of the gradients $\kappa_{(kl)}$ we find the following $6N$ equations:

$$[\kappa_{(\alpha\beta)}]_K^{K+1} = -[\Delta_{(\alpha\beta)}]_K^{K+1} \quad (5.14)$$

$$[c^{ijkl}\kappa_{(kl)}]_K^{K+1} n_j = 0 \quad (5.15)$$

$$\sum_{K=1}^N H_K c^{ijkl} \kappa_{(kl)} = 0 \quad (5.16)$$

The equation 5.10 gives us, in particular

$$\kappa_{(\alpha j)} n^j = 0 \quad (5.17)$$

and, hence, the formula of the equilibrium rotation can be simplified to read

$$[\kappa_{[ij]}]_K^{K+1} = [\Delta_{[\alpha\beta]}]_K^{K+1} \delta_i^\alpha \delta_j^\beta \quad (5.18)$$

According to the equation 5.18 the equilibrium rotation is generated by the rotation in the misfit only, and not by the equilibrium deformations.

5.4. Misfit stresses in the "free" γ/γ plate-like periodic structure with isotropic moduli

In the case of the γ/γ lamellar structure the elastic moduli of the lamellae occur to be equal, and the master system 5.15-5.16 can be simplified significantly to read:

$$c^{ijkl} [\kappa_{(kl)}]_K^{K+1} n_j = 0 \quad (5.19)$$

$$\sum_{K=1}^N H_K \kappa_{(kl)} = 0 \quad (5.20)$$

Using the equations 5.14, 5.19, 5.20 we find the following formula of the equilibrium stresses p_L^{ij} within the L -th lamella of the γ/γ laminated structure for the case of isotropic lamellae:

$$p_L^{ij} = \frac{\mu}{H} \sum_{K=1}^N H_K \left(\frac{2\nu}{1-\nu} \delta^{kl} \delta^{ij} + 2\delta^{ik} \delta^{jl} - \frac{2\nu}{1-\nu} \delta^{kl} n^i n^j \right) \left(\Delta_{(kl)}^K - \Delta_{(kl)}^L \right) \quad (5.21)$$

where μ is the shear modulus of the substance. Simple analysis shows that for the case under consideration the in-plane stresses $p_L^{\alpha\beta}$ only (α and β run the values 1 and 2) can assume non-zero values. For those very components the formula 5.21 can be simplified to read

$$p_L^{\alpha\beta} = \frac{\mu}{H} \sum_{K=1}^N H_K \left(\frac{2\nu}{1-\nu} \delta^{\alpha\beta} \delta^{\gamma\delta} + 2\delta^{\alpha\gamma} \delta^{\beta\delta} \right) \left(\Delta_{(\gamma\delta)}^K - \Delta_{(\gamma\delta)}^L \right) \quad (5.22)$$

5.5. Misfit stresses in the "free" α/γ - periodic structure (different elastic moduli)

5.5.1. α/γ - periodic structure (2-layers, isotropic elastic lamellae)

When considering the case of different elastic moduli we use the general system 5.14 - 5.16. First, we consider an analytical solution for the case of the isotropic elastic moduli and a periodic cell consisting of two lamellae: the one having the α -like structure and the other having the γ -like structure. A routine calculation gives the following result for the stresses within the lamellae:

$$\begin{aligned} p_1^{\alpha\beta} &= R_p \frac{2 \mu \mu_H}{\mu_H + \mu_H} \frac{122}{22} \left(\Delta_2^\gamma - \Delta_1^\gamma \right) \delta^{\alpha\beta} + \frac{2 \mu \mu_H}{\mu_H + \mu_H} \frac{122}{22} \left(\Delta_2^{(\alpha\beta)} - \Delta_1^{(\alpha\beta)} \right) \\ p_2^{\alpha\beta} &= -R_p \frac{2 \mu \mu_H}{\mu_H + \mu_H} \frac{121}{22} \left(\Delta_2^\gamma - \Delta_1^\gamma \right) \delta^{\alpha\beta} - \frac{2 \mu \mu_H}{\mu_H + \mu_H} \frac{121}{22} \left(\Delta_2^{(\alpha\beta)} - \Delta_1^{(\alpha\beta)} \right) \end{aligned} \quad (5.23)$$

where the constant R is defined as

$$R_p = \frac{\mu_H \left(1 + 3 C_1 \right) C_2 + \mu_H \left(1 + 3 C_2 \right) C_1}{\mu_H \left(1 + 3 C_1 \right) \left(1 + C_2 \right) + \mu_H \left(1 + 3 C_2 \right) \left(1 + C_1 \right)}, \quad C_K \equiv \frac{\nu_M}{1 - 2 \frac{\nu}{M}} \quad (5.24)$$

The in-plane deformations within the lamellae $\kappa_M^{(\alpha\beta)}$ are equal to

$$\begin{aligned} \kappa_1^{(\alpha\beta)} &= R_\kappa \left(C_2 - C_1 \right) \left(\Delta_2^\gamma - \Delta_1^\gamma \right) \delta^{\alpha\beta} + \frac{\mu_H}{\mu_H + \mu_H} \frac{22}{22} \left(\Delta_2^{(\alpha\beta)} - \Delta_1^{(\alpha\beta)} \right) \\ \kappa_2^{(\alpha\beta)} &= -R_\kappa \left(C_2 - C_1 \right) \left(\Delta_2^\gamma - \Delta_1^\gamma \right) \delta^{\alpha\beta} - \frac{\mu_H}{\mu_H + \mu_H} \frac{11}{22} \left(\Delta_2^{(\alpha\beta)} - \Delta_1^{(\alpha\beta)} \right) \end{aligned} \quad (5.25)$$

where

$$R_\kappa = \frac{\mu \mu_H H}{\mu_H \left(1 + 3 C_1 \right) \left(1 + C_2 \right) + \mu_H \left(1 + 3 C_2 \right) \left(1 + C_1 \right)} \quad (5.26)$$

For the out-plane deformation components $\kappa_M^{(33)}$ we get the formulae

$$\kappa_1^{(33)} = - \frac{\left(1 + 3 C_2 \right) C_1 \mu_H}{\mu_H \left(1 + 3 C_1 \right) \left(1 + C_2 \right) + \mu_H \left(1 + 3 C_2 \right) \left(1 + C_1 \right)} \left(\Delta_2^\gamma - \Delta_1^\gamma \right) \quad (5.27)$$

$$\kappa_2^{(33)} = \frac{\left(1 + 3 \frac{C_1}{C_2}\right) \frac{C_2 \mu H}{2 \cdot 1 \cdot 1}}{\frac{\mu H}{1 \cdot 1} \left(1 + 3 \frac{C_1}{C_2}\right) \left(1 + \frac{C_2}{C_1}\right) + \frac{\mu H}{2 \cdot 2} \left(1 + 3 \frac{C_2}{C_1}\right) \left(1 + \frac{C_1}{C_2}\right)} \left(\Delta_2^\gamma - \Delta_1^\gamma\right)$$

All other components of the deformation appear to be zero.

If the Poisson ratios of the phases are equal - $\nu_1 = \nu_2 = \nu$ - the formulae 5.24 - 5.27 allow some simplifications; in particular, we get the following formula of R in this case:

$$R_p = \frac{\nu}{1 - \nu} \quad (5.28)$$

If, in addition, the shear moduli are equal - $\mu_1 = \mu_2 = \mu$ - the formulae 5.23 of the stresses $p_M^{\alpha\beta}$ reads:

$$\begin{aligned} p_1^{\alpha\beta} &= \frac{2\mu H_2}{H_1 + H_2} \frac{\nu}{1 - \nu} \left(\Delta_2^\gamma - \Delta_1^\gamma\right) \delta^{\alpha\beta} + \frac{2\mu H_2}{H_1 + H_2} \left(\Delta_2^{(\alpha\beta)} - \Delta_1^{(\alpha\beta)}\right) \\ p_2^{\alpha\beta} &= -\frac{2\mu H_1}{H_1 + H_2} \frac{\nu}{1 - \nu} \left(\Delta_2^\gamma - \Delta_1^\gamma\right) \delta^{\alpha\beta} - \frac{2\mu H_1}{H_1 + H_2} \left(\Delta_2^{(\alpha\beta)} - \Delta_1^{(\alpha\beta)}\right) \end{aligned} \quad (5.29)$$

The formulae 5.29 appear to be in agreement with the formulae 5.23 as they should.

5.5.2. α/γ - periodic structure (N-layers, anisotropic elastic lamellae)

The analysis of this most general case is based on the system 5.14 - 5.16. It is convenient to introduce the following matrices:

$$T_{\pi\omega}^{\alpha\beta} = \sum_{K=1}^{K=N} H_{KK} c_{L\gamma\delta\pi\omega}^{\alpha\beta\gamma\delta} s_{L\gamma\delta\pi\omega}, \quad R^{\alpha\beta} = \sum_{K=1}^{K=N} H_{KK} c_{K(\gamma\delta)}^{\alpha\beta\gamma\delta} \left(\Delta_{K(\gamma\delta)} - \Delta_{L(\gamma\delta)}\right) \quad (5.30)$$

where $s_{L\gamma\delta\pi\omega}$ are the "partial 1-2" part of the compliance tensor s_{Lijkl} ; $c_{L\gamma\delta\pi\omega}^{\alpha\beta\gamma\delta}$ is the inverse of $s_{L\gamma\delta\pi\omega}$, i.e., the tensor satisfying the relationship

$$c_{L\gamma\delta\pi\omega}^{\alpha\beta\gamma\delta} s_{L\gamma\delta\pi\omega} = \delta_{(\pi}^\alpha \delta_{\omega)}^\beta \quad (5.31)$$

In terms of these "material" tensors the stress tensor within the L -th lamella $p_L^{\alpha\beta}$ can be calculated with the of the following linear system of equations:

$$T_{\pi\omega}^{\alpha\beta} p_L^{\pi\omega} = R^{\alpha\beta} \quad (5.32)$$

We remind the reader that in the equations 5.30 - 5.32 the indices run the values 1 and 2 only (not 3!).

The above program can be fulfilled explicitly in the case of isotropic lamellae. Corresponding calculations give the following formula of the stresses within the L -th lamella of the spatial period including N lamellae:

$$p_L^{\alpha\beta} = \sum_{M=1}^N H_M \left[\frac{2 \nu_M \mu_M T - (1 - \nu_M) \mu_L F}{(1 - \nu_M) 2 T_L (T + F)} \delta^{\alpha\beta} \delta_{\pi\omega} + \frac{\mu_M}{T_L} \delta_{(\pi}^{\alpha} \delta_{\omega)}^{\beta} \right] \left(\Delta_M^{(\pi\omega)} - \Delta_L^{(\pi\omega)} \right) \quad (5.33)$$

where T_L and F_L are defined as follows:

$$T_L = \frac{1}{2} \sum_{K=1}^N H_K \frac{\mu_K}{\mu_L}, \quad F_L = \sum_{K=1}^N H_K \frac{\mu_K}{\mu_L} \frac{\nu_K - \nu_L}{(1 - \nu_K)(1 - \nu_L)} \quad (5.34)$$

5.6. Misfit stresses in the "on-substrate" periodic multilayers with different elastic moduli

The case of the "on-substrate" multilayer is of the primary interest in many applications, especially, those related to the deposition of thin films, though they are of a secondary interest in concerns the TiAl lamellar structure, and we pay much less attention to it in this report. In the "on-substrate" case the extremely thick substrate does not deform practically and it can be treated as if it is rigid. Because of this fact all the monolayers of the multilayer are "aware" of presence of the substrate only, and they "know" nothing about the existence of other monolayers. Therefore, we can rewrite the system 5.14, 5.15

$$\kappa_{M(\alpha\beta)} = - \Delta_M^{(\alpha\beta)} \quad (5.35)$$

$$c_{M(kl)}^{ijkl} \kappa_{M(kl)} n_j = - c_M^{ijkl} \Delta_{M(kl)} n_j \quad (5.36)$$

where if the "misfit" with respect to the lattice of the substrate $\Delta_{M(kl)}$. The system 5.35, 5.36 appears to be closed with respect to the unknown misfit deformations $\kappa_{M(kl)}$.

It shows, in particular, that the stresses do not depend of the lamellae's thicknesses. In the case of isotropic lamellae we arrive at the following misfit deformations and stresses:

$$\kappa_{M(\alpha 3)} = 0, \quad \kappa_{M(33)} = \frac{\nu_M}{1 - \nu_M} \Delta_M^{\alpha\alpha} \quad (5.37)$$

$$p_{M(\alpha\beta)} = -2\mu_M \left(\nu_M \frac{\Delta\gamma}{M} \delta_{\alpha\beta} + \Delta_{(\alpha\beta)} \right) \quad (5.38)$$

6. ROLE OF VACANCIES IN SLOW EVOLUTION OF INTERFACES

6.1. Introduction

The elastic misfit stresses in TiAl alloys assume huge level of the order of the product of the elastic moduli times misfit deformations. The elastic energy associated with these stresses is enormously large also. Hence, there appears a driving force for the rearrangement of the species leading to the evolution accompanied by the creation of dislocations, migration of vacancies, development of interfacial corrugations, etc... The goal of this Section is to develop a thermodynamic approach allowing one to study the possibility of stress relaxation by means of mass rearrangement in the vicinity of the interface between the lamellae. We assume that the rearrangement is accompanied by interfacial diffusion and migration of the vacancies. Since the full consideration of the coherent interface solid-solid includes lots of calculation and details we illustrate our results and approach by considering the much simpler but conceptually a very close problem of mass rearrangement in the vicinity of a stress-free interface.

Mass transport through the bulk of solids is produced by the migration of vacancies whereas near surfaces the migration of vacancies is accompanied by surface diffusion, vaporization-sublimation, etc. (see, for instance, Christian [5]). The simplest and most robust features of the mechanism of surface diffusion has been studied by Mullins [20]. It is widely accepted that this mechanism dominates over migration of vacancies near traction-free surfaces. The situation can change dramatically at the nano-scale (see, for instance, Lee and Bennet [19]). At this level the very applicability of common thermodynamics seems questionable. Since at the nano-scale it is difficult to distinguish between the bulk and the surface part of the solid it seems natural, at least, to withdraw existing macro- and micro-scale "obvious" pre-assumptions and to explore the mechanism of interaction between surface diffusion and vacancy migration.

6.2. The master system of slow evolution

In this note three dissipation mechanisms are taken into account: i) dissipation from the flow of vacancies at the surface, ii) dissipation from bulk diffusion of vacancies, iii) dissipation from surface diffusion. In addition to the bulk energy ϕ the Herring-like surface energy [14] is taken into account. We consider a slightly corrugated solid film of thickness H attached to a rigid substrate (see, Fig. 3). It is assumed that the vacancies can penetrate the bulk through the free interface Ξ of the film only. We denote the time by t , the spatial coordinates by x^k (the spatial - Latin - indices run 1,2,3, and summation over repeated indices is implied). $\chi(x^k, t)$ is the volume

concentration of vacancies and $Q^i(x^k, t)$ is the vector of the volume flux of vacancies. We denote the surface coordinates of the free surfaces by ξ^β (the surface - Greek - indices run 1,2); $c(\xi^\beta, t)$ and $q^\alpha(\xi^\beta, t)$ are the normal velocity of the free surface and the surface flux of the species, respectively. Simple kinematics (Mullins, 1957 [20]; Christian, 1975 [5]; Grinfeld, 1994 [12]; Berdichevsky et al. 1996 [4]) leads to the following equations of mass conservation within the bulk

$$\frac{\partial \chi}{\partial t} = \frac{\partial Q^i}{\partial x^i} \quad (6.1)$$

and at the free surface

$$c(1 - \chi) = Q^i n_i + \frac{\partial q^\alpha}{\partial \xi^\alpha} \quad (6.2)$$

where $n_i(\xi^\beta, t)$ is a unit normal to the surface. The first term on the right hand side of Eq. 6.2 describes the migration of the free surface from the flow of vacancies through the surface, while the second describes the migration of the surface from the surface diffusion.

In this "kinematic surrogate" of a complete thermodynamic theory we ignore the role of stresses and heat and suppose that the energy per unit volume ϕ is a function of the vacancy concentration χ , whereas the surface energy per unit area is a positive constant σ . The simplest "constitutive" equations of the volume and surface fluxes compatible with the requirement of energy dissipation read:

$$Q^i = D_b^{ij} \frac{\partial}{\partial x^j} \frac{\partial \phi}{\partial \chi} \quad (6.3)$$

$$q^\alpha = D_s^{\alpha\beta} \frac{\partial}{\partial \xi^\beta} \frac{\phi - \sigma \kappa}{1 - \chi} \quad (6.4)$$

where D_b^{ij} and $D_s^{\alpha\beta}$ are the bulk and surface positive symmetric diffusion tensors, respectively; κ is a doubled mean curvature of the surface. The simplest equation to describe the flow of vacancies $Q^i n_i$ through the free surface compatible with the principle of energy dissipation reads

$$Q^i n_i = -D_I \frac{\phi + (1 - \chi)\phi_\chi - \sigma \kappa}{1 - \chi} \quad (6.5)$$

where D_I is a positive constant characterizing the resistance of the solid substance to penetrating of vacancies; $\phi_\chi = \partial \phi / \partial \chi$. For the sake of simplicity we assume that no vacancies flow through the film-substrate interface Σ :

$$Q^i n_i = 0 \quad (6.6)$$

The equations 6.1 - 6.6 form a closed self-consistent nonlinear system of the differential equations and boundary conditions for describing slow evolution of deformable elastic solids with vacancies.

We denote by χ° the equilibrium concentration of vacancies within the film when it has a plane free surface and is exposed to vacuum. Hence, according to Eq. 6.5, χ° satisfies the equation

$$\phi(\chi^\circ) + (1 - \chi^\circ)\phi_\chi(\chi^\circ) = 0 \quad (6.7)$$

6.3. Rate of change of small disturbances

Let us assume now that the free surface Ξ is slightly corrugated. The evolution of this corrugation can be explored by linearizing the equations 6.1 - 6.7 in the vicinity of the uniform configuration with further consideration of the solutions of the form $\sim \exp(\eta t + i k_\alpha \xi^\alpha)$; here ξ^α are the Cartesian coordinates on the equilibrium plane surface, k_α is the in-plane wave-number of the corrugation, η is the growth-decay rate of the corrugations.

In the following we assume that the film is isotropic and contains a small concentration of vacancies $\chi^\circ \ll 1$. Hence, the diffusion tensors take on the form: $D_b^{ij} = D_b \delta^{ij}$, $D_s^{\alpha\beta} = D_s \delta^{\alpha\beta}$, where D_b and D_s are positive constants whereas δ^{ij} and $\delta^{\alpha\beta}$ are the spatial and surface Kronecker symbols. A routine computation results in the following dispersion equation for this case:

$$\left[D_b k \tanh \left(kH \sqrt{1 + \frac{\eta}{k^2 N D_b}} \right) + D_I \right] (\eta + D_s \sigma k^4) + D_b D_I \sigma k^3 = 0 \quad (6.8)$$

where $N = \partial^2 \phi(\chi^\circ) / \partial \chi^2$.

Eq. 6.8 shows that the real η are necessarily negative as they should be in the absence of destabilizing factors.

If $\eta / (k^2 N D_b) \ll 1$ we can derive from Eq. 6.8 the following dispersion relation:

$$\eta = -\sigma k^3 \frac{D_s D_b k^2 \tanh(kH) + D_I D_s k + D_b D_I}{D_b k \tanh(kH) + D_I} \quad (6.9)$$

Eq. 6.9 shows clearly that the three dissipation mechanisms are not additive but interact with each other. It is quite instructive to extract from Eq. 6.9 Mullins' k^4 -term by rewriting the equation in the form

$$\eta = -D_s \sigma k^4 \left(1 + \frac{1}{\frac{D_s k^2}{D_I} \tanh(kH) + \frac{D_s k}{D_b}} \right) \quad (6.10)$$

The second terms in the brackets gives a correction to the " k^4 -law" which might even dominate, depending on the magnitudes of the nondimensional numbers $D_s k^2 / D_I$, $D_s k / D_b$, kH . First of all the assumption $\eta / (k^2 N D_b) \ll 1$ gives for the Mullins mode the following short wave-length limitation: $D_s \sigma k^2 / N D_b \ll 1$. The correction term approaches zero at $D_s k^2 / D_I \gg 1$ or $D_s k / D_b \gg 1$. For thin films the asymptote $kH \rightarrow 0$ is often of interest. In this limit Eq. 6.10 reads

$$\eta = -D_s \sigma k^4 \left(1 + \frac{D_b}{D_s k} \right) \quad (6.11)$$

and for sufficiently long wave corrugations with $k^{-1} \geq D_s/D_b$ the correction term is comparable with the Mullins k^4 -term. The k^3 -term appears to dominate as $k \rightarrow 0$ and in this limit the rate η approaches $-D_b\sigma k^3$. The k^3 variation also holds for thick films in the case of "instantaneous kinetics" (very large values of D_I). We notice that the same dispersion equation $\eta \propto k^3$ is valid for the transport mechanism "vaporization-condensation" which is presumably more effective than the vacancy transport at macro- and micro-scale level but the situation could be opposite at the nano-scale level because the values of D_s , D_b , and D_I need not be the same as they are in macroscopic bodies.

It is well-known that stresses within the solid can destabilize flat interfaces when rearrangement of mass is allowed. Assuming that the film is isotropic and it is sufficiently thick we arrive the following generalization of Eq. 6.8:

$$\left(D_b k \sqrt{1 + \frac{\eta}{k^2 N D_b}} + D_i \right) \eta + \left[\sqrt{1 + \frac{\eta}{k^2 N D_b}} (D_k D_i + D_b D_s k^2) k^2 + D_i D_s k^3 \right] \left[\sigma k - (1 - \nu) \frac{T^2}{\mu} \right] = 0 \quad (6.12)$$

where ν is the Poisson ratio, T is the uniaxial stress. Eq. 6.12 can be presented in the following dimensionless form:

$$(\sqrt{1+R} + K_{ib})R + \left[\sqrt{1+R}(K_{ib} + K_{sb}) + K_{ib}K_{sb} \right] [1 - (1 - \nu)\Phi\alpha] \Upsilon = 0 \quad (6.13)$$

where the dimensionless numbers R , K_{ib} , K_{sb} , $\Phi\alpha$, Υ are defined as follows:

$$R = \frac{\eta}{k^2 N D_b}, \quad K_{ib} = \frac{D_i}{D_b k}, \quad K_{sb} = \frac{D_s k}{D_b}, \quad \Phi\alpha = \frac{T^2}{\sigma \mu k}, \quad \Upsilon = \frac{\sigma k}{N} \quad (6.14)$$

At $R \ll 1$ we get the following nondimensional dispersion equation:

$$R = \frac{K_{ib} + K_{sb} + K_{ib}K_{sb}}{1 + K_{ib}} [1 - (1 - \nu)\Phi\alpha] \Upsilon \quad (6.15)$$

which gives a consistent generalization of the dispersion equation of Asaro and Tiller (1972) [3], compatible also with the static stability criteria (see, for instance, Grinfeld, 1991 [11]; Nozières, 1991 [21]) though it differs from the criterion of Vasudev, Asaro and Tiller (1975) [25] exploring the same mechanism of rearrangement via the bulk diffusion in pre-stressed crystal. The possibility of stress driven "necking" of a rod caused by the migration of vacancies has been noticed by Larche and Cahn (1985) [17].

7. CONCLUSION

7.1. The main results of this project.

The elastic misfit stresses in TiAl alloys assume huge level of the order of the product of the elastic moduli times misfit deformations. The elastic energy associated with

these stresses is enormously large also. Hence, there appears a driving force for the rearrangement of the material elements leading to the evolution accompanied by the creation of dislocations, migration of vacancies, development of interfacial corrugations, etc...

The goal of this research project was two-fold: i) to develop a theoretical approach which permits a self-consistent calculations of misfit stresses in TiAl lamellar multilayers and like structures, and ii) to develop a thermodynamic approach allowing one to study the possibility of stress relaxation by means of mass rearrangement in the vicinity of the interface between the lamellae.

We have developed a general scheme and establish several simple general formulae permitting to evaluate numerically and to analyze qualitatively misfit stresses in TiAl multilayers. On the basis of the model of coherent interfaces we arrive to the conclusion that one have to carefully distinguish between the "on-substrate" misfit stresses, on the one hand, and the "free-edge" misfit stresses, on the other case. The former case is more relevant for the stresses appearing at different processes of deposition: the misfit deformations and strains in the "on-substrate" multilayers can be calculated with the help of the master system 5.35, 5.36 (for the general anisotropic case) and simple formulae 5.37, 5.38 for the case of isotropic monolayers. The "free-edge" misfit deformations and stresses can be calculated with the help of the master system of algebraic equations 5.14 - 5.16 for the arbitrary anisotropic lamellae with different moduli. Much more simple working formulae can be used in the case of isotropic lamellae:

- i - 5.21, 5.22 for the case of equal elastic moduli (N-layer periodic multilayer in width),
- ii - 5.23-5.29 for the case of different elastic moduli (2-layer period multilayer in width),
- iii - 5.33, 5.34 for the case of different elastic moduli (N-layer periodic multilayer in width).

When studying possible mechanisms of stress relaxation we assumed that the rearrangement is accompanied by interfacial diffusion and migration of the vacancies. Since the full consideration of the coherent interface solid-solid includes lots of calculation and details we illustrate our results and approach by considering the much simpler but conceptually a very close problem of mass rearrangement in the vicinity of a stress-free interface. Using one of the approaches of irreversible thermodynamics of solids we explored the slow evolution of corrugations at the free boundary of a crystalline substance. The central issues of this study are the roles and the interaction of the three dissipation mechanisms: i) the surface diffusion of the species, ii) the bulk diffusion of vacancies, and iii) the flow of vacancies within the solid from vacuum. For the flux of vacancies from vacuum we have postulated the constitutive equation 6.5 which can be easily justified on the basis of Onsager ideology (this means, basically, that the constitutive law for the flux should obey the linearity between the flux and

corresponding driving force, and it has to provide the decay of the total accumulated energy). The new dissipative material constant D_I appearing in this modelling should be determined by quantitative measurement of different effects relating to the penetration of vacancies within the solid from vacuum (which is nothing else but rearrangement of the species accompanied by producing new vacancies in the vicinity of the free surface). One of the effects of this sort is the evolution of surface corrugations: the growth-decay rate of the corrugations depends, besides other parameters, on the value of D_I . We have derived a rather simple dispersion equation 6.8 of the growth rate of the surface corrugation η with the help of the master system of slow evolution (Eqs. 6.1 - 6.6) using the simplest model of the isotropic elastic solid with vacancies (Lifshitz, 1962 [18]). The equation 6.8 and its implications 6.9 - 6.11 allow one to find those corrugations for which the Mullins' k^4 - law gives acceptable approximation and those for which it does not. The equations 6.8 - 6.10 clearly show that the three above mentioned dissipative mechanism cannot be described as the linear superposition even for small disturbances (as they are treated usually in the existing literature).

As the side result of this study we have derived the dispersion equation 6.12 of the growth rate of surface corrugations in the pre-stressed crystal. Our dispersion equation appears to be consistent with the known thermodynamic criterion of the stress driven rearrangement instability as it should.

7.2. The most promising directions of future research

The result of this project demands further studies in what concerns both theory and experiment. The following directions of theoretical studies look the most urgent and promising:

I Misfit stresses in the case of large misfit deformations (actually, only the "in-plane" misfit deformations can be supposed small geometrically - of the order of several % - in the first approximation; the shear deformations are actually extremely big of the order of hundred %%).

II Misfit stresses in the plate-like random structure

III The edge effects

- The edge effects in vicinity of free surface
- The edge effects in vicinity of intersection of two free surfaces
- The edge effects at the corners

IV The grain structure of the lamellae

- An isolated grain within a lamella
- An equilibrium shape of an isolated grain within a lamella

- Stability of an isolated grain within a lamella
 - A periodic ensemble of the grains within the isolated lamella
- V Effective elastic moduli of laminated elastic structures with large internal stresses
- Effective elastic moduli in the γ/γ - periodic structure (equal elastic moduli)
 - Effective elastic moduli in the α/γ - periodic structure (different elastic moduli)
- VI The non-elastic effects
- Relaxation of misfit stresses via the mechanism of the stress driven rearrangement instability at the coherent interfaces
 - Relaxation of misfit stresses via proliferation of misfit dislocations
 - Nucleation of misfit dislocations via the mechanism of the stress driven rearrangement instability

References

- [1] P. Anderson: Scripta Met. Mat., 1992, vol. 27, pp. 687-692.
- [2] F. Appel, U. Christoph and R. Wagner: MRS Symposium 1994, vol. 318, pp. 691-696.
- [3] Asaro R. J. and Tiller W.A., *Metall. Trans.* **3** (1972) 1789.
- [4] Berdichevsky V.L., Hazzledine P.M. and Shoyhet B.I., (1996), submitted.
- [5] Christian J. W., Transformations in Metals and Alloys, (Pergamon Press), 1975.
- [6] Dimiduk D.M., Miracle Y.W. and Kim Y.W., *ISIJ Internat.* **31** (1991) 367.
- [7] Dimiduk D.M. and Parthasarathy T.A., *Phil. Mag. Lett.* **71** (1995), 21.
- [8] D. M. Dimiduk, Y. Q. Sun and P. M. Hazzledine: 1995, MRS Symposium vol. 364, pp. 599-605.
- [9] F. C. Frank and J. van der Merwe: Proc. Roy. Soc. A, 1949, vol. 198, pp.205-215.
- [10] L. B. Freund: J. Crystal Growth, 1993, vol. 132, pp. 341-344.
- [11] Grinfeld M.A., *Thermodynamic Methods in the Theory of Heterogeneous Systems*, (London: Longman) 1991.
- [12] Grinfeld M.A., *Scanning Microscopy*, **8** (1994) 869.
- [13] Hazzledine P.M., et al *MRS Symposium* **273** (1992) 81.

- [14] Herring C., In *Structure and Properties of Solid Surfaces*, edited by R. Gomer and C.S. Smith (University of Chicago Press) 1953.
- [15] Kad B.K. and Hazzledine P.M., *Phil. Mag. Lett.* **66** (1992), 133.
- [16] Kassner K. and Misbah C., *Europhys. Lett.* **28** (1994) 245.
- [17] Larché F.C. and Cahn J.W., *Acta Metall.* **33** (1985) 331.
- [18] Lifshitz I.M., *Soviet Physics, ZETP* **17** (1962) 909.
- [19] Lee M.Y. and Bennet P.A., *Phys. Rev. Lett.* **75** (1995) 4460.
- [20] Mullins W.W., *J. Appl. Phys.* **28** (1957) 333.
- [21] Nozières Ph., Lectures at Beg-Rohu Summer School (1989), in *Solids Far From Equilibrium*, edited by C. Godrèche, (Cambridge: Cambridge Univ. Press) 1991.
- [22] T. Pfullmann and P. A. Beaven: *Scripta Met. Mat.*, 1993, vol. 28, pp. 275-280.
- [23] M. Riemer, H. Biermann and H. Mughrabi: 1996, Private communication.
- [24] Shoykhet B.A. Private communication.
- [25] Vasudev P., Asaro R.J. and Tiller, A.W., *Acta Metall.* **23** (1975) 341.
- [26] Y. Umakoshi, T. Nakano and T. Yamane: *Mat. Sci. Engnr. A*, 1992, vol. 152, pp. 81- 87.

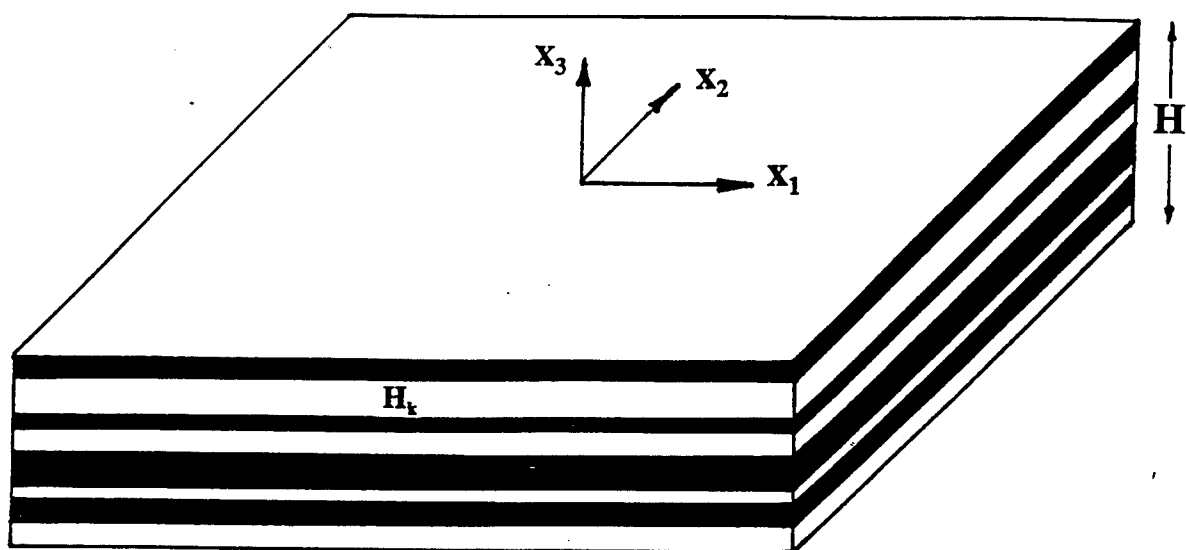


Figure 1.

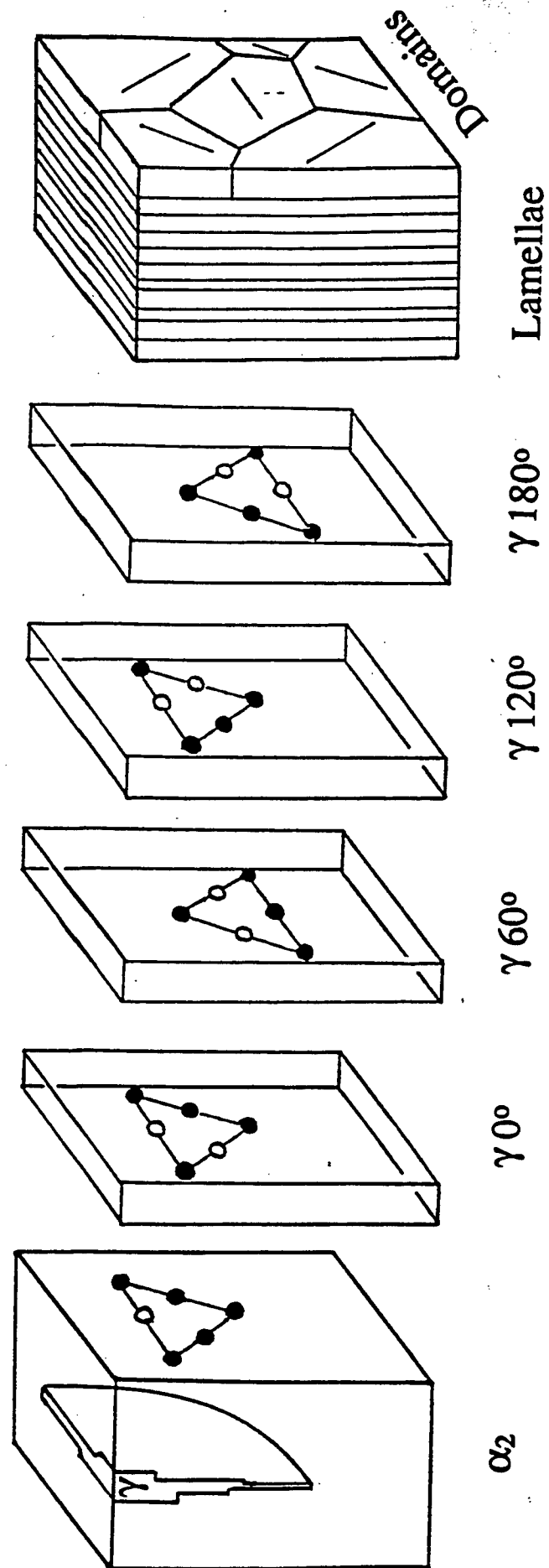


Figure 2.

INTERFACIAL SLIDING IN BRITTLE FIBROUS COMPOSITES

Michael C. Larson, Assistant Professor
Department of Mechanical Engineering

Tulane University
New Orleans, LA 70118-5674

Final Report for:
Summer Research Extension Program
Wright Materials Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Ceramics Group
Wright Materials Laboratory

December 1997

INTERFACIAL SLIDING IN BRITTLE FIBROUS COMPOSITES

Michael C. Larson, Assistant Professor
Department of Mechanical Engineering
Tulane University

Abstract

The extent of interfacial sliding is pivotal to the mechanical behavior of ceramic matrix composites. This research illuminates the relative importance of interfacial friction, interfacial toughness, and interfacial roughness in determining the sliding resulting from an impinging matrix crack. A computational model based on a two dimensional surface integral formulation was used to model interfacial behavior. During the project period, a novel experiment was developed for cracks at frictional, rough interfaces which shows the three-dimensional nature of the crack-interface interaction. The experiments, coupled with the modeling, can provide guidance for tailoring the crack resistance of composites in structural applications.

INTERFACIAL SLIDING IN BRITTLE FIBROUS COMPOSITES

Michael C. Larson

1.0 INTRODUCTION

Research into ceramic matrix composites has sprung from the desire for strong, tough materials at extremely high temperatures (1000-1500 degrees Celsius). Monolithic ceramics are inherently brittle. Oxide-oxide material systems are promising for introducing a measure of crack resistance while retaining desirable material properties. In these materials, the fiber/matrix interface plays the key role in controlling the global toughness of the composite. The degree of interfacial toughness, roughness, and friction will decide the length of the interfacial sliding zones. The size of these sliding zones will provide a measure of pseudo-ductility by controlling the matrix crack spacing, and thus, the global toughness of the material.

The objective of this work is to link a mechanics-based computational modeling technique with a materials-based experimental program aimed at identifying and optimizing toughening mechanisms in ceramic composites. The authors have been engaged in applying the surface integral method to model fractures at interfaces between brittle constituents; researchers at the Wright Materials Laboratory have shown through their experiments that interfaces play the key role in determining the crack resistance in technologically important brittle matrix composites. By coupling these two, we are seeking to refine our mechanistic representation of the interfaces in order to predict composite performance and to design optimal interfaces.

2.0 BACKGROUND AND MOTIVATION

In brittle composites, the length of the interfacial sliding zone at the periphery of a matrix fracture plays an essential role in determining the global toughness of the composite. This is manifest in two ways: the sliding relieves the stress concentration which is transferred to the fibers and hence regulates fiber failure, and the sliding influences matrix fracture spacing which in turn dictates energy dissipation.

Previous research by the author employed the double cleavage drilled compression specimen (DCDC) introduced by Janssen to observe the behavior of a crack as it proceeds toward and past a frictional fiber/matrix interface. As the DCDC specimen is compressed, a stable fracture sweeps down its mid-plane (Warren, 1987). The specimens were modified to include fibers oriented perpendicular to the plane of fracture. Due to the brittle nature of glass, the assumption of local crack extension at a constant fracture toughness is valid and any variation of the crack front is assumed to be due to the inserted fibers. The study revealed the optimal levels of friction for friction-only interfaces so that the material's resistance to fracture is maximized. Crack profiles are displayed in Figure 1 for three cases of a specimen as is, one with holes at 1.905 cm away, and one with fibers at 1.905 cm away. This shows the significant crack retardation that can be achieved from the insertion of fibers.

Figure 1. Crack fronts taken at constant time intervals for DCDC specimens with a) no hole b) hole and no fiber and c) fiber

That work suggested that the influence of the interface in controlling the extent of the interfacial sliding could be characterized by the local angle between the three-dimensional crack front and the surface of the fiber. This work pursued these two notions, computationally determining the combinations of interfacial

properties that will yield desired interfacial sliding lengths and experimentally measuring the crack angle as a function of the friction, toughness, and roughness of the interface.

3.0 SOLUTION METHODOLOGY

3.1 Computational Method

A two-dimensional surface integral method, similar to the displacement discontinuity method described by Crouch (1976), is used to investigate the behavior of a crack intersecting a planar interface possessing toughness, roughness, and friction. The necessary singular fundamental solutions are derived from the elasticity solution of a point force in an infinite medium (Wiles and Curran, 1982).

The total relative displacement at any point of the cracked surfaces is obtained by summing the effects of a continuous distribution of the fundamental solutions over the entire cracked domain. This procedure leads to the governing singular integral equation

$$\sigma_{ij}(\mathbf{x}) = \int \frac{\partial}{\partial \zeta} \Gamma_{ijk}(\mathbf{x}, \zeta) \delta_k(\zeta) dA \quad , \quad (\text{EQ 1})$$

where σ_{ij} are the stresses at a field point, \mathbf{x} is the location of the field point, Γ_{ijk} are the kernel solutions, δ_k are the crack displacement, and ζ is a vector which spans the fracture surfaces (all referenced to a fixed Cartesian frame).

The approximate numerical solution to the governing integral equation is obtained by transforming it to a system of algebraic equations with the nodal displacements as unknown quantities. The opening and sliding displacements are approximated over discrete elements of the fracture surface by interpolation functions which depend upon values at specified nodal points. The crack domain is divided into E discrete elements, i.e.,

$$\sigma_{ij}(\mathbf{x}) = \sum_{e=1}^E \int \frac{\partial}{\partial \zeta} \Gamma_{ijk}(\mathbf{x}, \zeta) \delta_k^{(e)}(\zeta) dA \quad . \quad (\text{EQ 2})$$

Crack displacements, δ , are calculated by boundary collocation. Interpolation functions relate an assumed displacement variation to the nodal values:

$$\delta_k^{(e)}(\xi) = \sum_{n=1}^N h(\xi)^{(n)} \hat{\delta}_k^{(n)} \quad . \quad (\text{EQ } 3)$$

Constant displacement interpolations are used for the majority of the elements which compose the fractures in order to reduce the computational time needed per simulation. Elements with specialized high order interpolations are necessary, however, for near-tip regions having large displacement gradients. This higher order interpolation is based upon the first two terms of the asymptotic elastic crack displacement field governed by the stress intensity factor, K :

$$\delta_i = K_i 4 \frac{(1-\nu^2)}{E\sqrt{2\pi}} \rho^{1/2} + B_i(\nu, E, n) \rho^{3/2} \quad , \quad (\text{EQ } 4)$$

with ρ being the distance from the crack front and the subscript i is not a summation index but rather corresponds to the fracture mode, $i = \text{I or II}$.

Gauss quadrature is used to numerically integrate the distribution of Green's functions for the two-dimensional displacement discontinuities in a homogeneous, isotropic, linear elastic material. For frictional sliding, the normal stress is compared to the shear local to each element. If the local normal traction is tensile, then the crack elements open and are free to slide; if the current local normal traction is compressive, and the shearing traction exceeds the normal traction multiplied by a friction coefficient, then the element is closed and slides against a restraining shearing traction. The slip length is defined as the length that the crack has traveled up the interface.

3.2 Experimental Method

The goal of this ongoing research effort is to measure experimentally how different interfacial surface roughness, coupled with variable friction and bonding, act to limit the crack opening displacement of an impinging fracture. These measurements are compared with computational predictions derived from the

two-dimensional modeling scheme described above. The measurements bolster confidence in using the computational tools to give design guidance for interfaces. In addition, the experiments reveal the steady-state angle at which a three dimensional crack front propagates when it intersects a rough, frictional interface. This quantity is useful as a parameter to embed into three dimensional computational simulations where there are often numerical problems associated with surfaces.

The dual double cleavage drilled compression (dual-DCDC) specimens used in this experiment, a variation on the geometry used by Janssen (1974), were fabricated from plate glass and are shown schematically in Figure 3 on page 9. The dimensions of both the drilled and undrilled halves were 2.54 cm X 10.16 cm X 0.9525 cm. The crack starter holes, drilled in the center of the 0.9525 cm X 10.16 cm faces, have a diameter of 0.3175 cm. A half containing a hole was butted up against a non-drilled specimen with the sides in contact possessing a known roughness. The pre-stress normal to the interface was controlled using a hydraulic piston assembly that pressed the two specimens together. The dual-DCDCs were compressed in an special apparatus designed specifically for these tests, which is shown in Figure 2.

As the specimen is compressed, a crack initiates at the drilled hole due to the stress concentration located there. The stress concentration is similar to that of a remote stress applied to a hole in an infinite sheet (Warren, 1987). Due to the specimen geometry, a stable crack can be propagated perpendicular to the controlled interface. Because of the brittle nature of glass, we assume that local crack extension is at a constant fracture toughness and any variation in the crack front shape as it traverses the specimen is a consequence of the roughness and friction at the interface.

While the cracks are propagating, the crack opening displacements are determined using crack opening interferometry (COI) (Liechti, 1994). COI is based on the concept that when light is sent through two opposing surfaces the reflected beams interfere with one another creating interference fringes along the specimen. These fringes can then be used to determine the amount that a crack is opening or closing. One can represent the combination of the two beams by adding the two light vector magnitudes together where each light vector is represented by

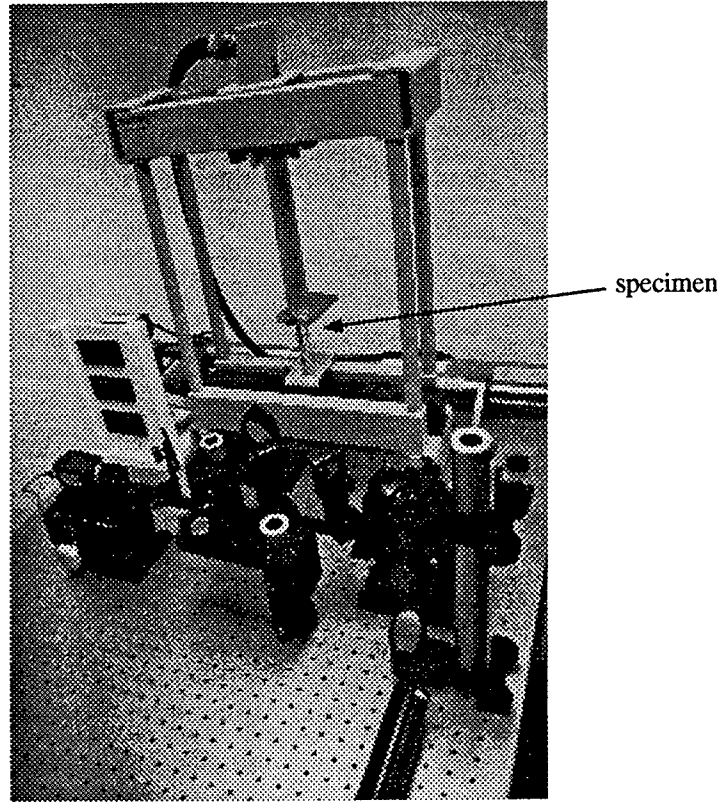


Figure 2. 100kN press on a vibration isolation table with optics for COI measurements.

$$E = a \cos(\phi - \omega t) \quad (\text{EQ } 5)$$

and their combination is

$$E_{total} = \bar{a} \cos\left(\frac{1}{2}(\phi_1 + \phi_2) - \omega t\right) \quad (\text{EQ } 6)$$

where

$$\bar{a} = a \sqrt{2[1 + \cos(\phi_2 - \phi_1)]} \quad (\text{EQ } 7)$$

is the amplitude of the resulting wave. The phase angle difference is related to the displacement δ by

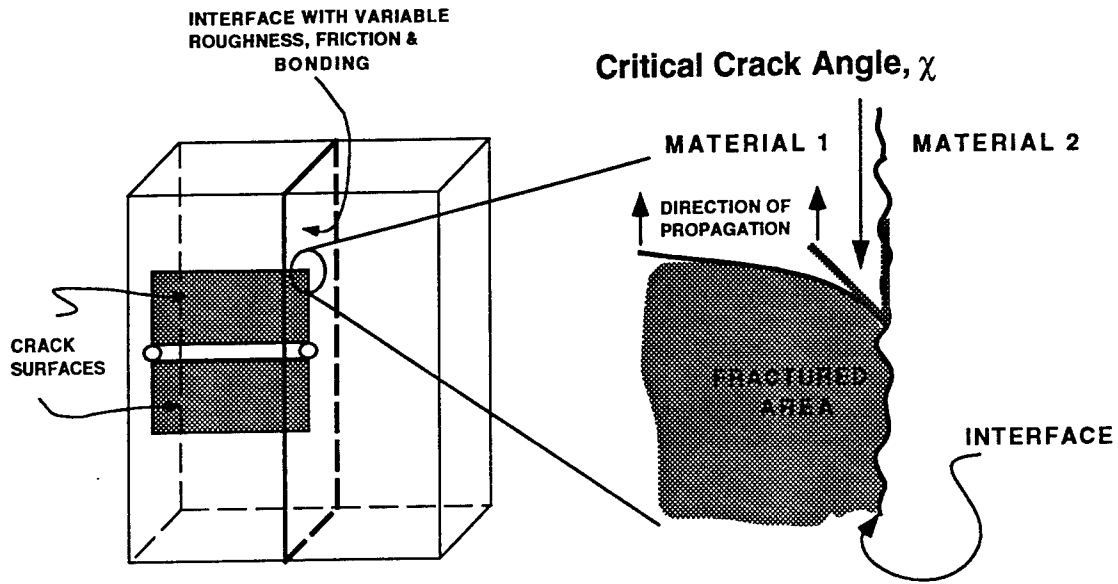


Figure 3. Dual DCDC specimen showing the crack bowing angle χ at the frictional, rough interface.

$$\phi_2 - \phi_1 = 2\pi \frac{\delta}{\lambda} \quad (\text{EQ 8})$$

Substituting into the previous equation and knowing that the intensity is proportional to the square of the amplitude, one gets

$$I \approx 4a^2 \left(\cos\left(\pi \frac{\delta}{\lambda}\right) \right)^2 \quad (\text{EQ 9})$$

It can be seen from this equation that the dark fringes occur at where the intensity is a minimum ($I=0$) and the light fringes occur where the intensity is at a maximum.

$$I_{min} = 0 \Rightarrow \delta = (2n+1) \frac{\lambda}{2} \quad (\text{EQ 10})$$

$$I_{max} \Rightarrow \delta = n\lambda \quad (\text{EQ 11})$$

Thus, destructive interference, or dark fringes, can be observed where

$$h = \frac{n\lambda}{2n_2 \cos \gamma} \quad (\text{EQ 12})$$

and constructive interference, or bright fringes, can be seen where

$$h = \frac{(2n - 1)\lambda}{4n_2 \cos \gamma} \quad (\text{EQ 13})$$

where n_2 is the refractive index of the fluid between the crack surfaces. For these experiments, the fluid is air and the refractive index is equal to 1.003. For a wavelength λ the resolution is equal to $1/4\lambda$.

The mode I stress intensity factor for any fringe can then be found from the normal crack opening displacements by

$$K_I = \frac{n\lambda}{16n_2 \cos \gamma} \left(\frac{E}{(1 - \nu^2)} \right) \left(\frac{2\pi}{r} \right)^{\frac{1}{2}} \quad (\text{EQ 14})$$

4.0 RESULTS

4.1 Computational Method

To reveal the near-tip behavior of interfacial sliding, that will not be influenced by the finite geometry of the matrix crack, we limit the absolute sliding length to be less than 8% of the half crack length. This rule of thumb was determined by keeping the difference between the numerically determined stress field around the crack tip within 5% of the analytical asymptotic stress field. The results, which are in agreement with those presented in Dollar and Steif (1989), for the relationship between the normalized slip length and the Coulomb coefficient of friction are shown in Figure 4.

Roughness is assumed to be approximated by a triangular arrangement of surface integral elements with a period, P , and amplitude, A . Good correlation was found in the results using the same roughness angle, ϕ , defined as

$$\phi = \operatorname{atan} \frac{2A}{P} \quad (\text{EQ 15})$$

as long as the period of roughness was no more than ten percent of the sliding length. One important note about rough interfaces when there is not bonding (i.e., no toughness) is that the sliding zones are not continuous, closed sliding, but rather are interspersed with open and sticking portions. Sample design

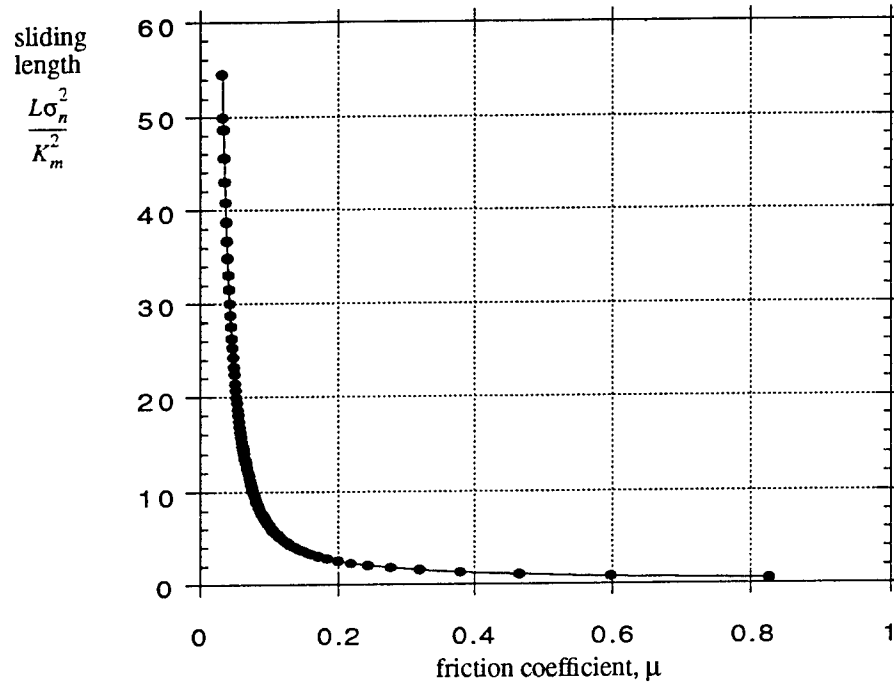


Figure 4. Effect of friction on sliding length of interfacial toughness.

surfaces showing sliding lengths as a function of toughness, friction, and roughness are displayed in Figure 5.

4.2 Experimental Method

All of the drilled halves of the dual-DCDC specimens were pre-cracked by loading to 1600 kg with pop-in occurring between 1200-1500 kg. For the three levels of roughness (polished, 300 μm , 1000 μm) at a constant friction coefficient (glass on glass) and no interfacial toughness, a range of lateral compression was applied (from 0-0.25 MPa). The crack bowed at a much lower lateral compression for the polished specimens than the abraded specimens. For all specimens that exhibited this crack pinning nature, a steady

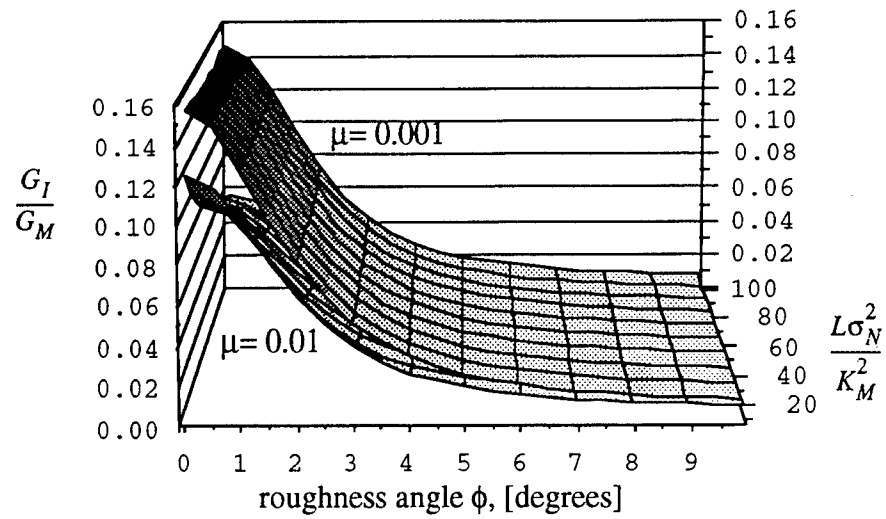


Figure 5. Interfacial sliding length variation for combinations of interfacial friction and roughness for interfaces with no toughness.

state angle between the crack front and the interface was observed. The higher the roughness, the smaller the angle between the crack front and interface was. The angle is dependent on the amount of lateral compression. Crack kinking was also observed if the specimens were put under a Mode II as well as a Mode I loading. These specimens exhibited growth out of plane and a significant pinning at the interface.

	25 kPa	30 kPa	100 kPa	250 kPa
Polished	63	62	58	
300 μm		66	46	
1000 μm		67	67	37

TABLE 1. Average crack angles, χ , as a function of the nominal normal pressure at the interface and the roughness.

5.0 CONCLUSIONS

The design criteria for sliding length, e.g. Figure 5, demonstrate that designing for a particular sliding length can be virtually impossible for low friction ($\mu < 0.01$) and high roughness (> 5 degrees) interfaces. The sliding length is very sensitive to variations in small values of friction coefficient and it is unreasonable to pinpoint a particular length. At high friction and roughness, the interfacial sliding is inhibited for all but the lowest of interfacial bonding. For moderate ranges of energy release rate, coefficient of friction, and roughness angle, tailoring the sliding length is a reasonable objective. The most control comes at moderate friction with low roughness.

The experiments are continuing to yield valuable information concerning the influence that the interface has on the character of the matrix crack growth. The stable crack growth achieved in our dual-DCDC geometry used in the test rig we have constructed, coupled with the optical measuring system we have put in place, will make it practical to evaluate different interface types to simulate those being considered in practical ceramic matrix composites. This information will also be useful in guiding the development of three dimensional computer models for near-interface cracks.

6.0 REFERENCES

- Brebbia, C.A., and Dominguez, J. (1989), *Boundary Elements An Introductory Course*, McGraw-Hill, New York.
- Carter, W. C., E. P. Butler, and E. R. Fuller (1991), Micro-mechanical aspects of asperity-controlled friction in fiber-toughened ceramic composites, *Scripta Met.* 25, 579.
- Crouch, S. L. (1976), Solution of plane elasticity problems by the displacement discontinuity method, *Int. J. Num. Meth. Eng.* 10, 301.
- Curtin, W. A. (1991), Theory of mechanical properties of ceramic matrix composites, *J. Am. Ceram. Soc.* 74 (11), 2837.
- Dollar, A. and P.S. Steif (1989), A tension crack impinging upon frictional interfaces, *J. Appl. Mech.* 56, 291.
- Evans, A.G. (1990), Perspective on the development of high-toughness ceramics, *J. Am. Ceram. Soc.* 73(2), 187.
- Fares, N. (1989), Crack fronts trapped by arrays of obstacles: numerical solutions based on surface integral representation, *J. Appl. Mech.* 56, 837.
- He, M.Y. and J.W. Hutchinson (1989), Crack deflection at an interface between dissimilar elastic materials, *Int. J. Solids Struct.* 25 (9), 1053.
- He, M.Y., J.W. Hutchinson, and A.G. Evans (1994), Crack deflection at an interface between dissimilar elastic materials: role of residual stresses, *Int. J. Solids Struct.* 31 (24), 3443.

- Herbel, R. (1995), An Experimental Investigation of Frictional Interface Effects on Toughening in Composites, Master's Thesis, Tulane University.
- Janssen, C. (1974), Specimen for Fracture Mechanics Studies on Glass, *Proceedings of the Tenth International Congress on Glass*, The Ceramic Society of Japan, 23-30.
- Jero, P. D., and R. J. Kerans (1991), The Contribution of Interfacial Roughness to Sliding Friction of Ceramic Fibers in a Glass Matrix, *Scripta Metallurgica et Materialia*, **24**, 2315-18.
- Jero, P. D., R. J. Kerans, and T. A. Parthasarathy (1991), Effect of interfacial roughness on the frictional stress measured using pushout tests, *J. Am. Ceram. Soc.* **74** (11), 2793.
- Keat, W.D., B.S. Annigeri, , and M.P. Cleary (1988), Surface integral and finite element hybrid method for two- and three-dimensional fracture mechanics analysis, *Int. J. Fract.* **36**, 35.
- Kerans, R. J., R. S. Hay, N. J. Pagano, and T. A. Parthasarathy (1989), The Role of the Fiber-Matrix Interface in Ceramic Composites, *Ceramic Bulletin*, **68** [2] 429-41.
- Kerans, R. J. and T. A. Parthasarathy (1991), Theoretical analysis of the fiber pullout and pushout tests, *J. Am. Ceram. Soc.* **74** (7), 1585.
- Larson, M.C. (1995), Fracture propagation near a frictionally-constrained fiber interface, *Int. J. Composites Eng.* **5**, 25.
- Larson, M.C. (1997), Comparison of experimental and computational models for three dimensional crack/fiber interactions, *Exper. Mech.* **35** (4), in press.
- Leichti, K. M. (1994), On The Use of Classical Interferometry Techniques in Fracture Mechanics, in *Experimental Techniques in Fracture*, 95-124.
- Liu, H.-Y., Zhou, L.-M., and Mai, Y.-W. (1995), Effect of Interface Roughness on Fiber Push-Out Stress, *Journal of the American Ceramic Society*, **78** [3], 560-66.
- Love, A.E.H. (1927), *A treatise on the mathematical theory of elasticity*, Cambridge University Press, Cambridge.
- Mackin, T. J., Yang, J., and Warren, P. D. (1992), Influence of Fiber Roughness on the Sliding Behavior of Sapphire Fibers in TiAl and Glass Matrices, *Journal of the American Ceramic Society*, **75** [12], 3358-62.
- Mackin, T. J., P. D. Warren, and A. G. Evans (1992), Effects of fiber roughness on interface sliding in composites, *Acta Met.*, **40** (6), 1251.
- Parthasarathy, T.A., D. R. Barlage, P. D. Jero and R. J. Kerans (1994), Effect of interfacial roughness parameters on the fiber pushout behavior of a model composite, *J. Am. Ceram. Soc.* **77** (12), 3236.
- Parthasarathy, T.A. and R.J. Kerans (1996), Anticipated effects of interfacial roughness on the behavior of selected ceramic composites, in press.
- Rabinowicz, Ernest (1995), *Friction and Wear of Materials*, John Wiley and Sons, Inc., New York.
- Rice, J.R. (1988), Elastic fracture mechanics concepts for interfacial cracks, *J. Appl. Mech.* **55**, 98.
- Warren, W. E. (1987), Theoretical Analysis of the Double Cleavage Drilled Compression Specimen, *International Journal of Fracture* **33** 223-35.
- Wiles, T.D. and J.H. Curran (1982), A general 3-d displacement discontinuity method, *Proceedings of the 4th International Conference on Numerical Methods in Geomechanics*, Vol. 1, Edmonton, Canada, 103.

**TOOLS FOR THE ANALYSIS AND DESIGN
OF GAIN SCHEDULED MISSILE AUTOPILOTS**

**Douglas A. Lawrence
Associate Professor
School of Electrical Engineering
and Computer Science**

**Ohio University
Athens, Ohio 45701**

Final Report for:

**Summer Research Extension Program
Wright Laboratory
Armament Directorate
Eglin Air Force Base**

Sponsored by:

**Air Force Office of Scientific Research
Bolling Air Force Base, DC**

and

**Wright Laboratory
Armament Directorate
Eglin Air Force Base**

December 1997

TOOLS FOR THE ANALYSIS AND DESIGN OF GAIN SCHEDULED MISSILE AUTOPILOTS

Douglas A. Lawrence
Associate Professor
School of Electrical Engineering
and Computer Science
Ohio University

Abstract

The development of computational and analytical tools for the analysis and design of gain scheduled missile autopilots is summarized. *Matlab*TM routines for general-purpose multivariate splining and numerical solution of a class of partial differential equation are described. The splining routines are used to create a *Simulink*TM simulation and a routine to extract linearized models for a nonlinear six degree-of-freedom air-to-air missile model. Also, a stability analysis is presented that lends justification to the potentially hazardous yet commonly practiced technique of designing linear autopilots on the basis of linearizations about non-equilibrium conditions.

TOOLS FOR THE ANALYSIS AND DESIGN OF GAIN SCHEDULED MISSILE AUTOPILOTS

Douglas A. Lawrence

1. Introduction

Gain scheduling techniques originated in the design of flight control systems. Since the aerodynamic characteristics of an aircraft or missile can vary drastically over a specified flight envelope, a linear controller designed on the basis of a linearized aerodynamic model about a single flight condition will typically fail to deliver adequate global performance. In practice, this difficulty is remedied by performing linear controller designs at a set of flight conditions that spans the flight envelope and captures, in some sense, the nonlinear aerodynamic effects. The parameters, often simply gains, in the resulting set of linear controllers are then interpolated or scheduled as a function of a selected set of scheduling variables.

Recently, gain scheduling has received a great deal of attention in the literature as a formal control system design method. The PI has collaborated on the development of a particular approach to gain scheduling that emphasizes nonlinear systems aspects and is described in broad terms by the follows steps: [6]

- determine a nonlinear plant's family of constant operating points or equilibria, parameterized by a designated set of scheduling variables;
- for the corresponding family of linearized plants, design a family of linear controllers to meet prescribed design goals at each constant operating point;
- construct a gain scheduled controller that linearizes to the appropriate linear controller at each constant operating point;
- check nonlocal performance of the resulting nonlinear control system.

The emphasis in this approach is on the linearization requirement imposed on the gain scheduled controller in the third step. Specifically, technical conditions have been derived for the existence of a gain scheduled controller with this property. Failure to satisfy these conditions results in the introduction of so-called *hidden coupling terms* that can potentially degrade system performance.

The PI's participation in the 1996 AFOSR Summer Faculty Research Program involved the application of this gain scheduling methodology to the design of a pitch channel autopilot for a hypothetical air-to-air missile model. [5] The work described in this report builds on these efforts and is divided into two main activities:

- develop practical tools for applying this gain scheduling methodology to missile autopilot synthesis;

- analyze the stability mechanisms at work in existing missile autopilot implementations.

Each of these topics is described in detail in the following two sections. The final section contains some concluding remarks.

2. Tools for Gain Scheduled Autopilot Synthesis

In general, the following computations must be performed in order to implement the gain scheduling methodology outlined above:

- solve systems of nonlinear algebraic equations in order to characterize the plant's family of constant operating points;
- calculate partial derivatives in order to form the family of linearized plant models;
- perform parameterized linear controller design based on the family of linearized plant models;
- solve partial differential equations to construct a gain scheduled controller that meets the linearization requirement described above.

In realistic problems, these steps can rarely be carried out analytically. This is the case for missile autopilot design problems as even though the equations of motion for a missile have an analytical structure in terms of the various aerodynamic coefficients, the coefficients themselves are usually represented in tabular form. At the outset, this necessitates a numerical approach to the above computations. Towards developing a set of computational tools that can be systematically applied in a general missile autopilot design setting, *Matlab*TM code has been developed to perform the following tasks:

- construct multivariate splines from tabular data;
- evaluate partial derivatives of these multivariate splines;
- generate approximate solutions to a type of partial differential equation that arises in the implementation gain scheduled controllers that satisfy the previously described linearization requirements.

In addition, these general purpose routines have been used to construct a *Simulink*TM simulation of a six degree-of-freedom air-to-air missile model provided by the Navigation and Control Branch, Armament Directorate, Wright Laboratory (WL/MNAG), Eglin Air Force Base. Also for this missile model, a *Matlab*TM routine has been developed that calculates a linearized model about a user-selected operating point. In the remainder of this section, an overview of these general purpose *Matlab*TM routines along with the *Simulink*TM missile simulation and linearization routine is presented.

Bivariate splining can be implemented in the *Matlab*TM workspace using the *bs_make* and *bs_evalp* commands that correspond to function m-files of the same name. The *bs_make* command is used to create a bivariate spline from a two-dimensional data table. The resulting spline is evaluated

using the *bs_evalp* command. Partial derivatives through order three in each variable can also be evaluated by specifying optional arguments. On-line help messages describing proper syntax for each of these commands are shown in Figures 2.1 and 2.2.

```
» help bs_make

Make a bivariate spline matrix to be used with bs_evalp.m

Usage: [a]=bs_make(x,y,z)

Inputs: x - vector of break points in x variable
        y - vector of break points in y variable
        z - matrix of function values with rows corresponding to
            different values of x, columns corresponding to different
            values of y

Output: a - matrix containing spline data to be used with bs_evalp.m

Steven Andrew
10-16-97 Version 1.2

»
```

Fig. 2.1: Matlab on-line help message for the *bs_make* command

```
» help bs_evalp

Evaluate a bivariate spline at a point (x,y)

Usage: [f]=bs_evalp(a,x,y[,x_der,y_der])

Inputs: a    - bivariate spline matrix created using bs_make.m
        x    - x coordinate of evaluation point
        y    - y coordinate of evaluation point
        x_der - partial order (1, 2, or 3) w.r.t. x variable (optional)
        y_der - partial order (1, 2, or 3) w.r.t. y variable (optional)

Output: f    - spline value at point (x,y) or partial at (x,y)
              when used with x_der and y_der

Steven Andrew
2-17-97 Version 1.1

»
```

Fig. 2.2: Matlab on-line help message for the *bs_evalp* command

In an analogous fashion, trivariate splining can be implemented in the *Matlab*TM workspace using the *ts_make* and *ts_evalp* commands. The *ts_make* command is used to create a trivariate spline from a three-dimensional data table organized into a two-dimensional array in a particular fashion. The resulting spline is evaluated using the *ts_evalp* command. Partial derivatives through order three in each variable can also be evaluated by specifying optional arguments. On-line help messages describing proper syntax for each of these commands are shown in Figures 2.3 and 2.4.

```

» help ts_make

Make a trivariate spline matrix to be used with ts_evalp.m

Usage:  [a]=ts_make(x,y,z,f)

Inputs: x - vector of break points in x variable
        y - vector of break points in y variable
        z - vector of break points in z variable

        f - matrix of data points with initial n rows corresponding to
            different values of x and m columns corresponding to different
            values of y. Values of z are represented by o nxm blocks
            stacked in a column matrix - forming a matrix with
            (n x o) rows and m columns

Output: a - matrix containing spline data to be used with ts_evalp.m

Steven Andrew
10-16-97  Version 1.2

»

```

Fig. 2.3: *Matlab* on-line help message for the *ts_make* command

The implementation of gain scheduled controllers that satisfy the previously described linearization requirement often involves functions of the scheduling variables that are solutions to a certain type of partial differential equation. Specifically, with Θ a q -dimensional vector of scheduling variables, a function $x: R^q \rightarrow R$ is sought that satisfies

$$\frac{\partial x}{\partial \Theta}(\Theta) = G(\Theta) \quad (2.1)$$

where $\partial x / \partial \Theta = [\partial x / \partial \theta_1 \quad \dots \quad \partial x / \partial \theta_q]$ is the gradient of x and $G(\Theta) = [G_1(\Theta) \quad \dots \quad G_q(\Theta)]$ has been determined from the preceding steps in the gain scheduling design process. A solution to (2.1) exists if and only if $G(\Theta)$ satisfies the mixed-partial conditions

```

» help ts_evalp

Evaluate a trivariate spline at a point (x,y,z)

Usage: [f]=ts_evalp(a,x,y,z[,x_der,y_der,z_der])

Inputs: a      - trivariate spline matrix created using ts_make.m
        x      - x coordinate of evaluation point
        y      - y coordinate of evaluation point
        z      - z coordinate of evaluation point
        x_der  - partial order (1, 2, or 3) w.r.t. x variable (optional)
        y_der  - partial order (1, 2, or 3) w.r.t. y variable (optional)
        z_der  - partial order (1, 2, or 3) w.r.t. z variable (optional)

Output: f      - spline value at point (x,y,z) or partial at (x,y,z)
                  when used with x_der, y_der, and z_der

Steven Andrew
2-17-97  Version 1.1
»

```

Fig. 2.4: Matlab on-line help message for the *ts_evalp* command

$$\frac{\partial G_i}{\partial \Theta_j} = \frac{\partial G_j}{\partial \Theta_i}, \quad 1 \leq i < j \leq q. \quad (2.2)$$

Since typically $G(\Theta)$ will not satisfy these restrictive conditions, an exact solution to (2.1) will not exist and it is therefore of interest to pursue approximate solutions.

One approach is to use techniques from finite element analysis, in particular, collocation methods. This will next be described for the case $q = 2$. Let

$$\Gamma = \left\{ \Theta = (\Theta_1, \Theta_2) \mid \Theta_1^{\min} \leq \Theta_1 \leq \Theta_1^{\max}, \Theta_2^{\min} \leq \Theta_2 \leq \Theta_2^{\max} \right\}$$

denote a rectangular domain in R^2 on which an approximate solution to (2.1) is desired. A strictly increasing sequence in each variable, $\Theta_i^{\min} = \Theta_i^1 < \dots < \Theta_i^{n_i} = \Theta_i^{\max}$, $i = 1, 2$, defines a grid on the domain Γ .

The goal is to construct a bivariate spline, denoted $\hat{x}(\Theta_1, \Theta_2)$, that has the required partial derivative values at each grid point. That is,

$$\frac{\partial \hat{x}}{\partial \Theta_k}(\Theta_1^i, \Theta_2^j) = G_k(\Theta_1^i, \Theta_2^j), \quad i = 1, \dots, n_1, \quad j = 1, \dots, n_2, \quad k = 1, 2 \quad (2.3)$$

Using a bivariate spline that is cubic in each variable also allows $\hat{x}(\Theta_1, \Theta_2)$ to have continuous second order partial derivatives. Further, spline values at any two grid points may be freely selected.

Calculating and subsequently evaluating a bivariate spline with these properties can be accomplished in the *Matlab*[™] workspace using the commands *fem2d* and *eval2d*. On-line help messages describing proper syntax for each of these commands are shown in Figures 2.5 and 2.6.

```
»help fem2d

Implement 2-dimensional finite element method
to construct a bivariate spline that represents
an approximate solution to a type of partial
differential equation

Usage: [a] = fem2d(x, y, p, q, const1, const2)

Inputs: x, y - vectors containing grid values in each variable
        p    - partial derivative values w.r.t. x at grids points
        q    - partial derivative values w.r.t. y at grids points
        const1 - spline value at (x(1), y(1))
        const2 - spline value at (x(n), y(1))

Output: a    - matrix containing spline data to be used with eval2d.m

Haoyu Lai
11-5-97

»
```

Fig 2.5: *Matlab* on-line help message for the *fem2d* command

```
» » help eval2d

Evaluate the bivariate spline constructed by fem2d.m at a point (xe, ye)

Usage: [val] = eval2d(a, x, y, xe, ye)

Inputs: a    - matrix containing spline data generated by fem2d.m
        x, y - vectors containing grid values in each variable
        xe, ye - coordinates of evaluation point

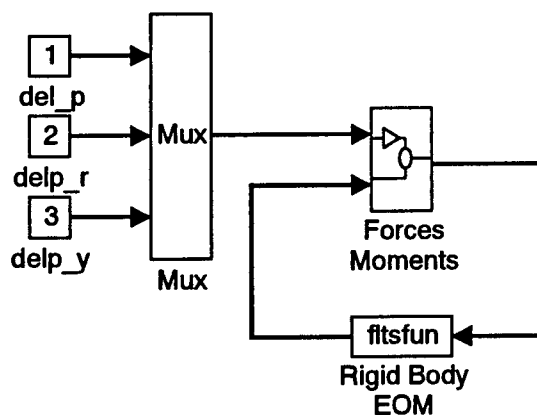
Output: val  - value of the spline at the point (xe, ye)

Haoyu Lai
7-25-97

»
```

Fig 2.6: *Matlab* on-line help message for the *eval2d* command

One of the objectives of this work was to convert the previously mentioned six degree-of-freedom air-to-air missile simulation from *Fortran* code to a *Simulink*[™] model in which the aerodynamic coefficients are implemented using bivariate and trivariate splines. The top level *Simulink*[™] block diagram for the simulation is shown in Figure 2.7. The aerodynamics are divided into two blocks labeled *Forces/Moments* and *Rigid Body EOM*. The *Forces/Moments* subsystem block has two vector inputs and one vector output. The first (top) input is a three-dimensional vector consisting of the effective pitch, roll, and yaw control surface deflections, all in radians. The second (bottom) input is a seven-dimensional vector consisting of angle-of-attack (radians), sideslip (radians), missile velocity (meters/sec), roll, pitch, and yaw body rates all in radians/sec, and altitude in meters. The output is a six dimensional vector containing the three components of the aerodynamic force vector (Newtons) and the three components of the aerodynamic moment vector (Newton-meters), all relative to the standard body-fixed coordinate system. The *Rigid Body EOM* s-function block has a vector input corresponding to the output of the *Forces/Moments* subsystem block and a vector output corresponding to the second (bottom) input of the *Forces/Moments* subsystem block. The *Rigid Body EOM* s-function block implements the standard differential equations describing the motion of a rigid body under the influence of external forces and moments. Gravitational forces are calculated within this block so it is not necessary to pass the Euler angles to the *Forces/Moments* subsystem block.



Six Degree-of-Freedom Air-to-Air Missile Simulation

Fig. 2.7: *Simulink* top-level block diagram for six degree-of-freedom air-to-air missile simulation

Simulation results are presented for the initial conditions given in Table 2.1 and a doublet pulse described by

$$\delta(t) = \begin{cases} 0 \text{ rad} & 0 \leq t < 0.1 \text{ sec} \\ -10 \frac{\pi}{180} \text{ rad} & 0.1 \leq t < 0.2 \text{ sec} \\ 10 \frac{\pi}{180} \text{ rad} & 0.2 \leq t < 0.3 \text{ sec} \\ 0 \text{ rad} & t \geq 0.3 \text{ sec} \end{cases}$$

applied sequentially to each of the three control surface deflections, with the other two fixed at zero deflection. Responses for a doublet pitch fin deflection are shown in Figures 2.8 through 2.10. Responses for a doublet roll fin deflection are shown in Figures 2.11 through 2.13. Responses for a doublet yaw fin deflection are shown in Figures 2.14 through 2.16. As an indication of the execution speed of the simulation, one second of simulation time on a 75 MHz Pentium PC with 16 MB RAM running *Matlab 4.2c* and *Simulink 2.0* took slightly less than two minutes for each of the three simulation scenarios.

Table 2.1: Initial Condition Values for *Simulink* Simulations

Variable	Initial Value	Units
angle-of-attack	0	radians
sideslip angle	0	radians
missile velocity	500	meters per second
roll angle	0	radians
pitch angle	0	radians
yaw angle	0	radians
roll rate	0	radians per second
pitch rate	0	radians per second
yaw rate	0	radians per second
altitude	10,000	meters

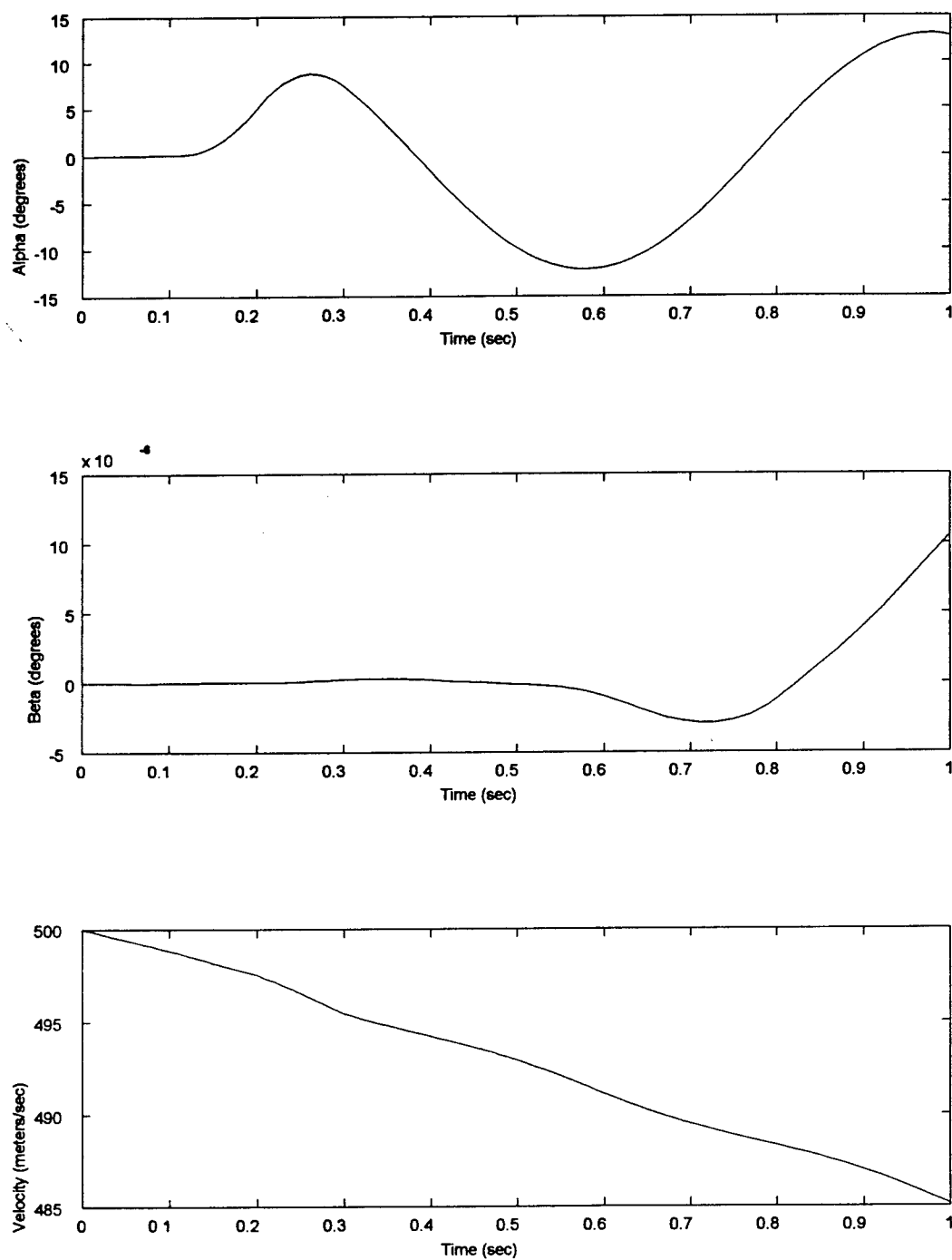


Fig 2.8: Angle-of-attack α (top), side-slip β (middle), and velocity (bottom) vs. time for doublet pitch fin deflection

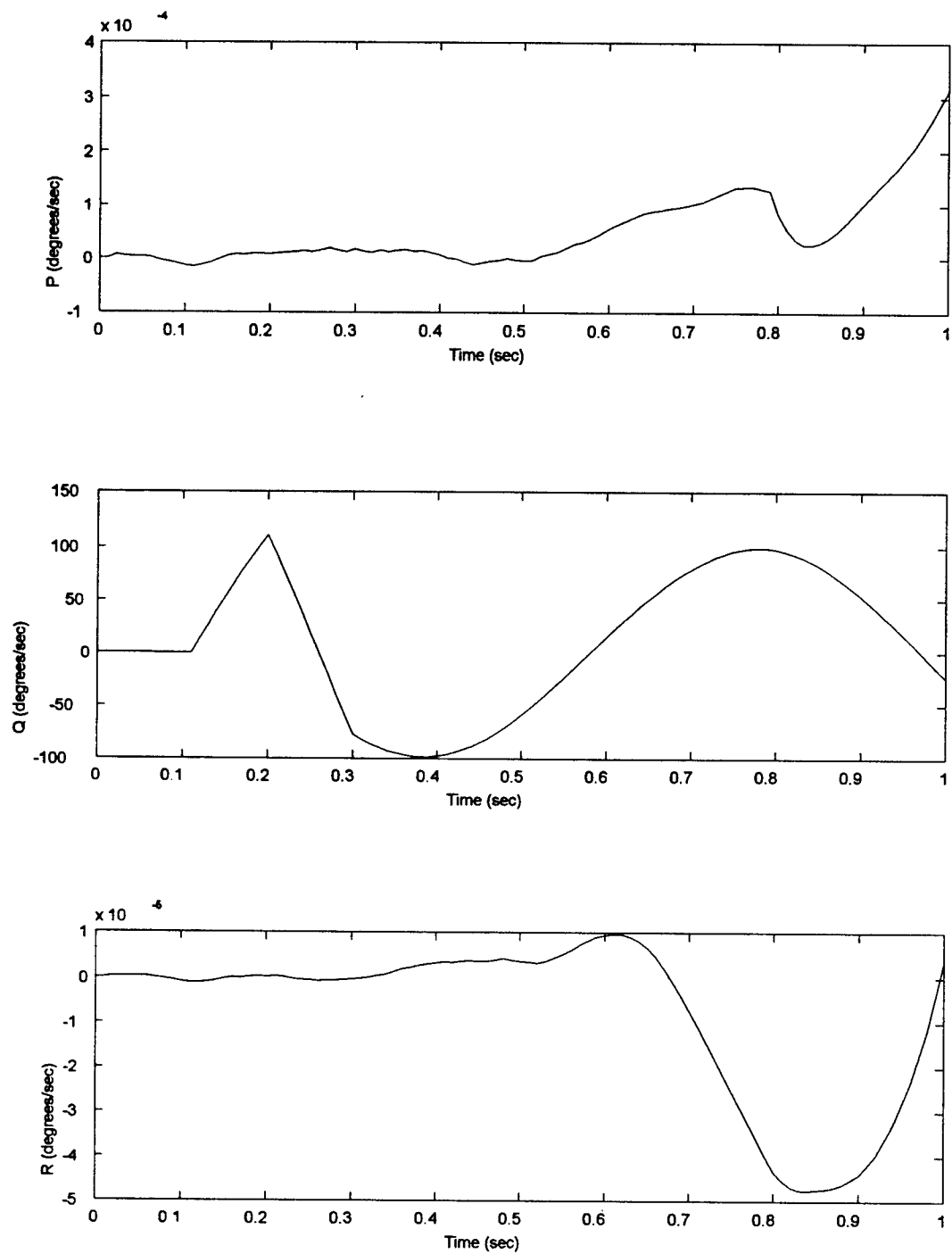


Fig 2.9: Roll rate P (top), pitch rate Q (middle), and yaw rate R (bottom) vs. time for doublet pitch fin deflection

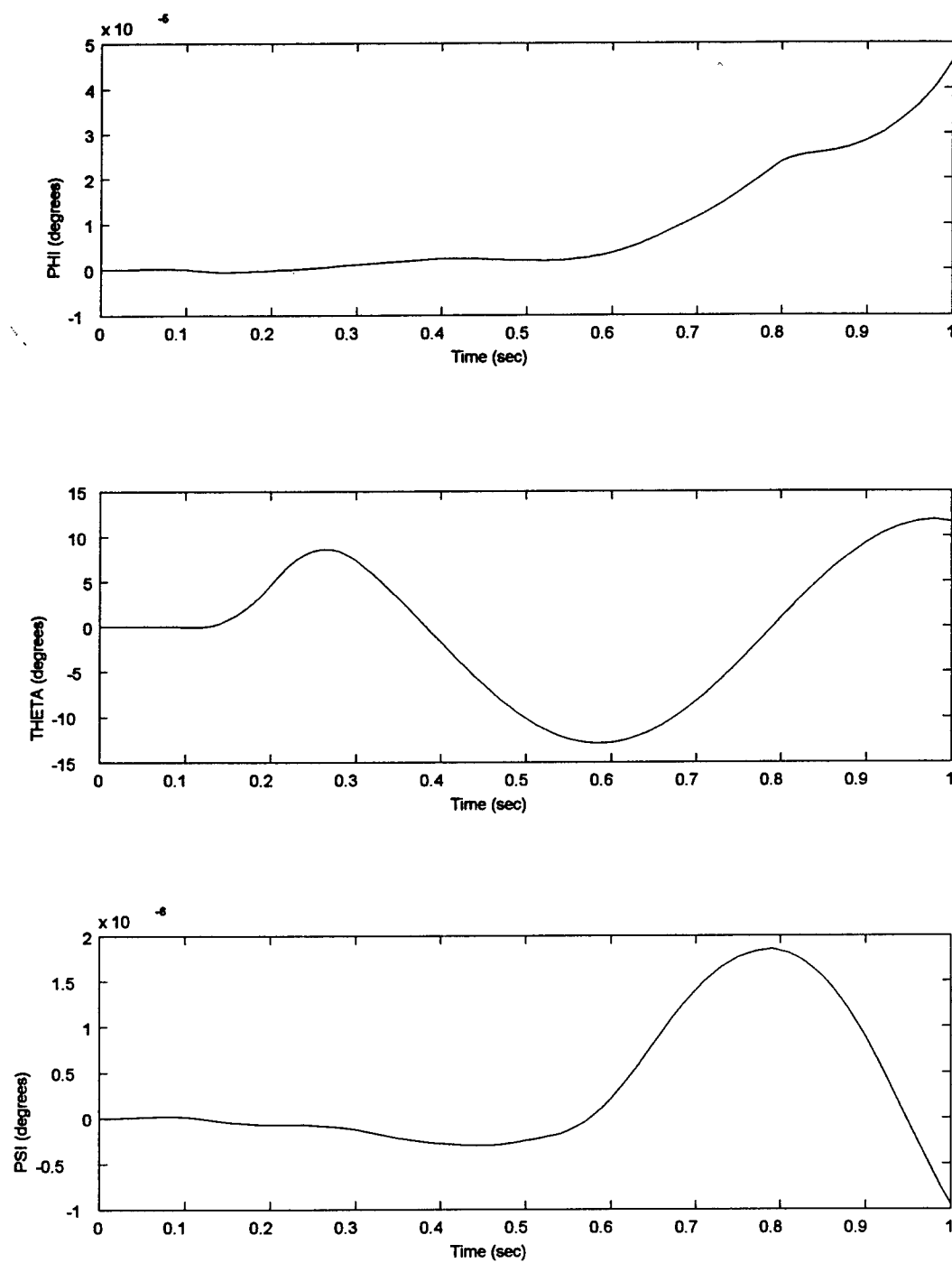


Fig 2.10: Euler angles ϕ (top), θ (middle), and ψ (bottom) vs. time for doublet pitch fin deflection

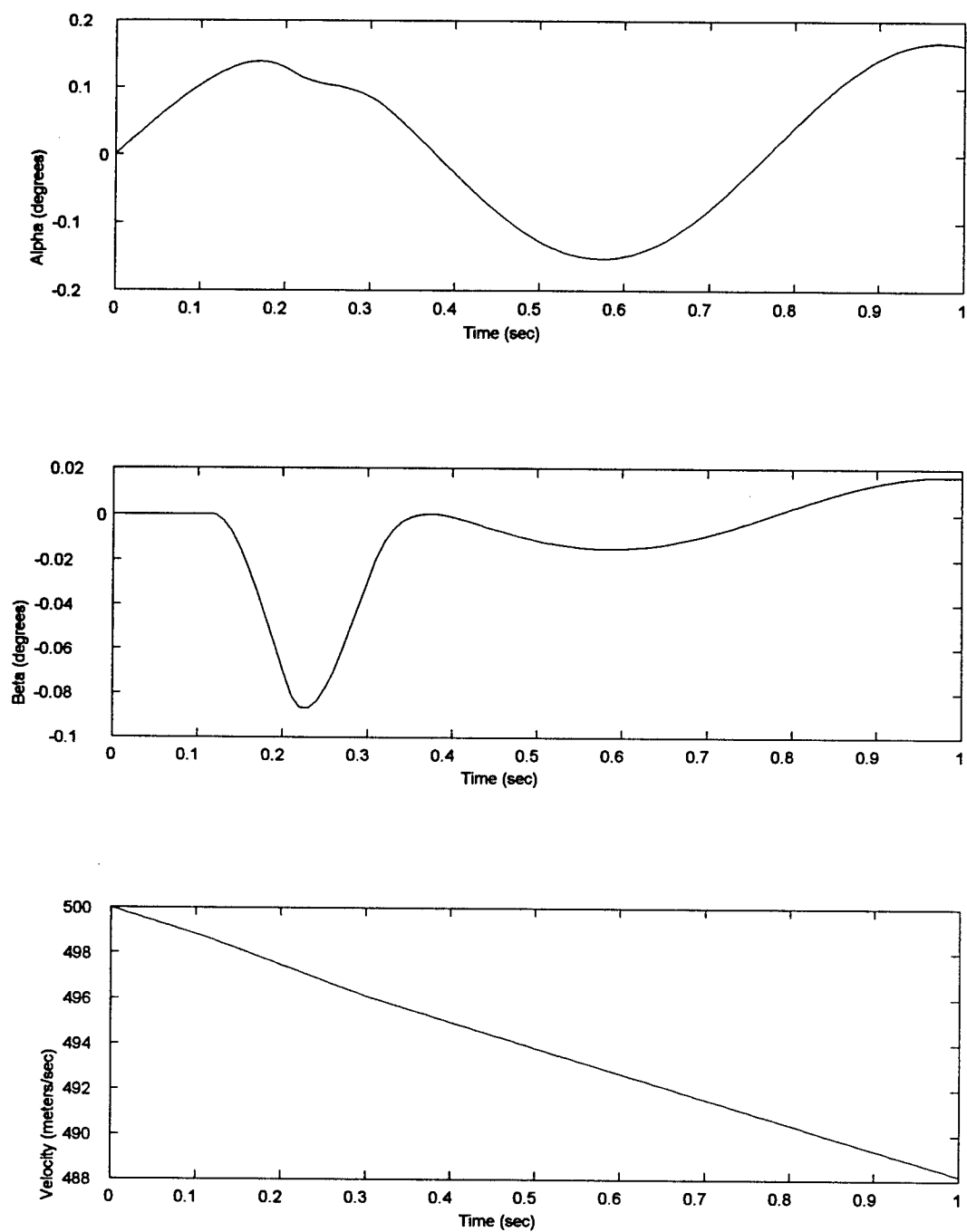


Fig 2.11: Angle-of-attack α (top), side-slip β (middle), and velocity (bottom) vs. time for doublet roll fin deflection

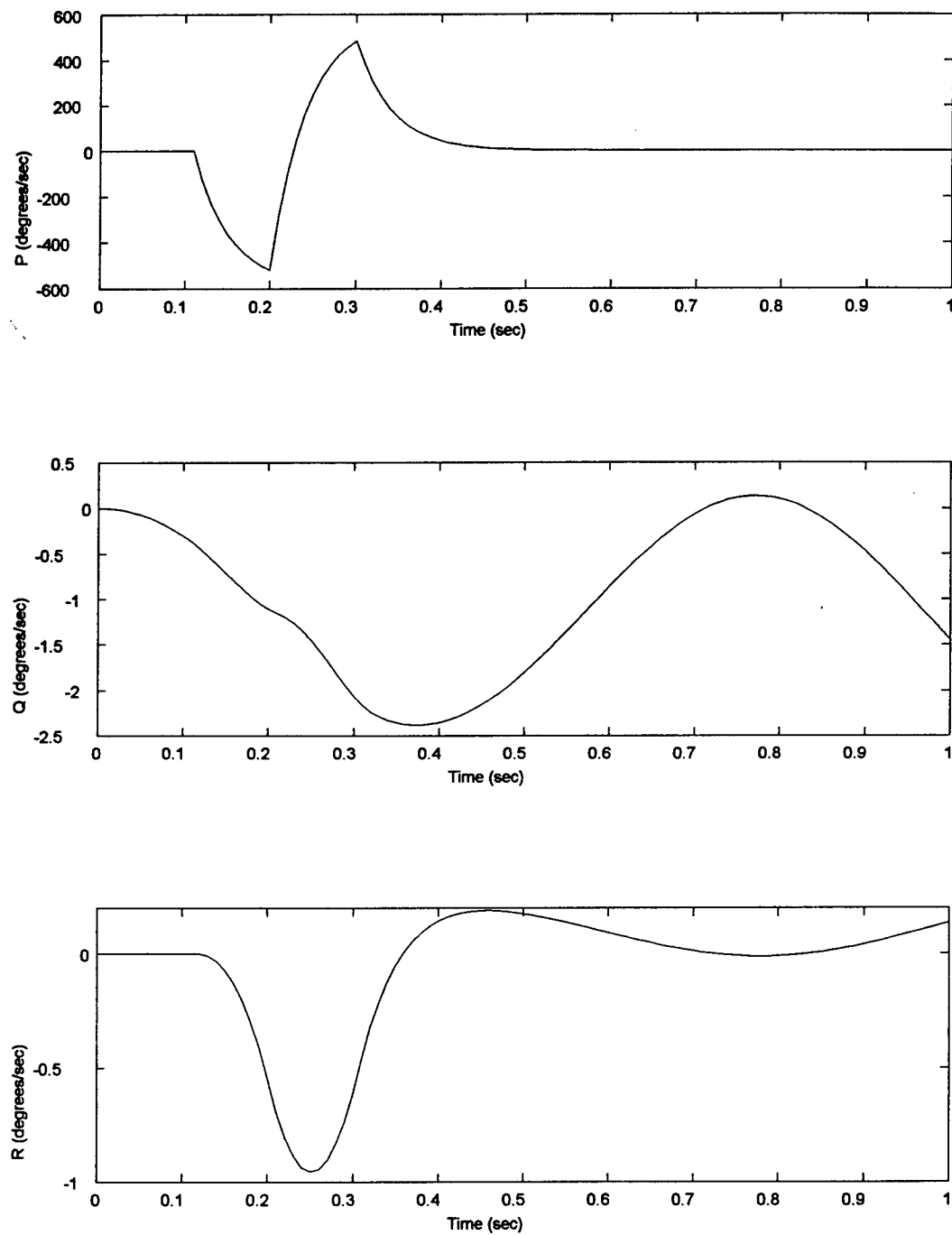


Fig 2.12: Roll rate P (top), pitch rate Q (middle), and yaw rate R (bottom) vs. time for doublet roll fin deflection

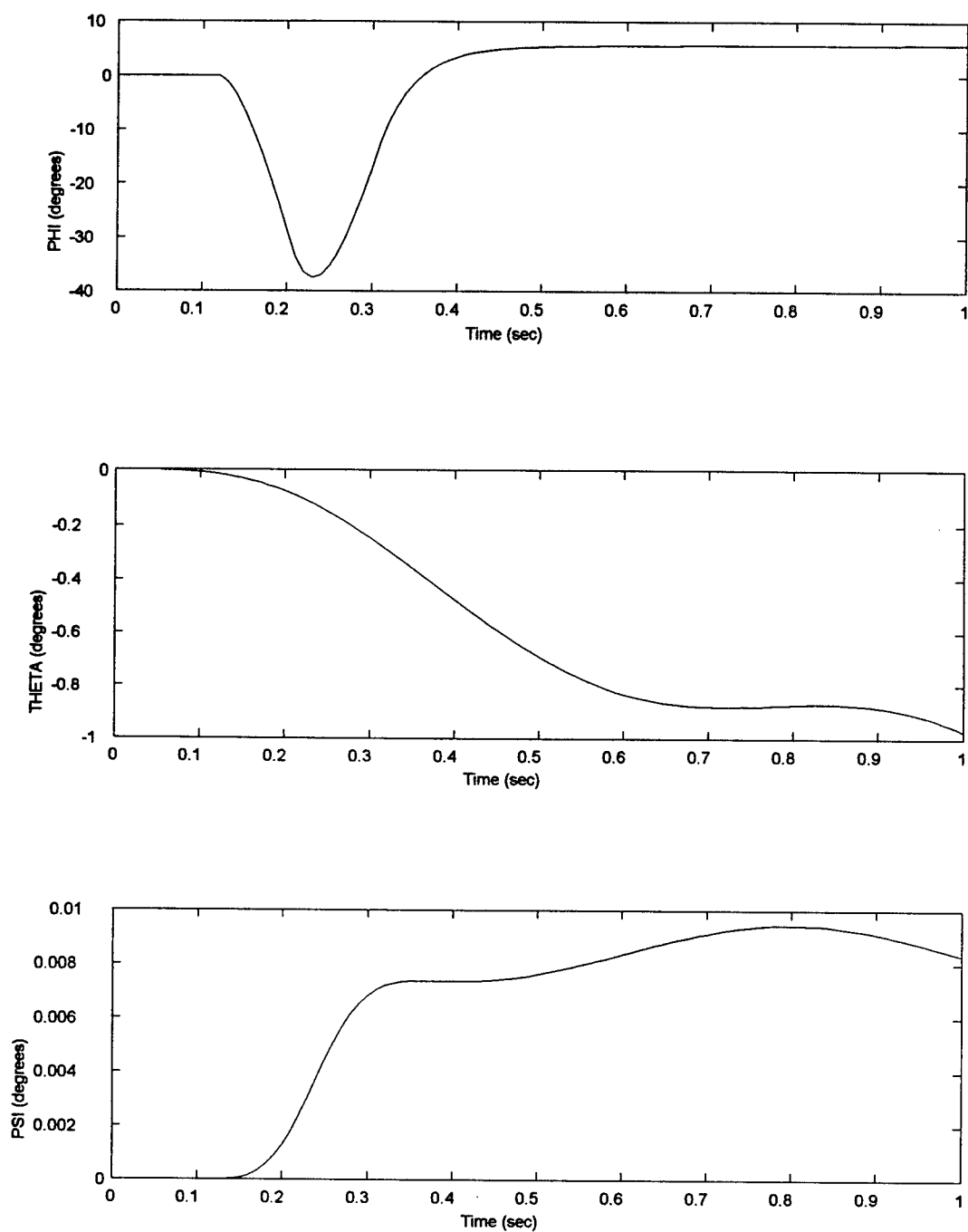


Fig 2.13: Euler angles ϕ (top), θ (middle), and ψ (bottom) vs. time for doublet roll fin deflection

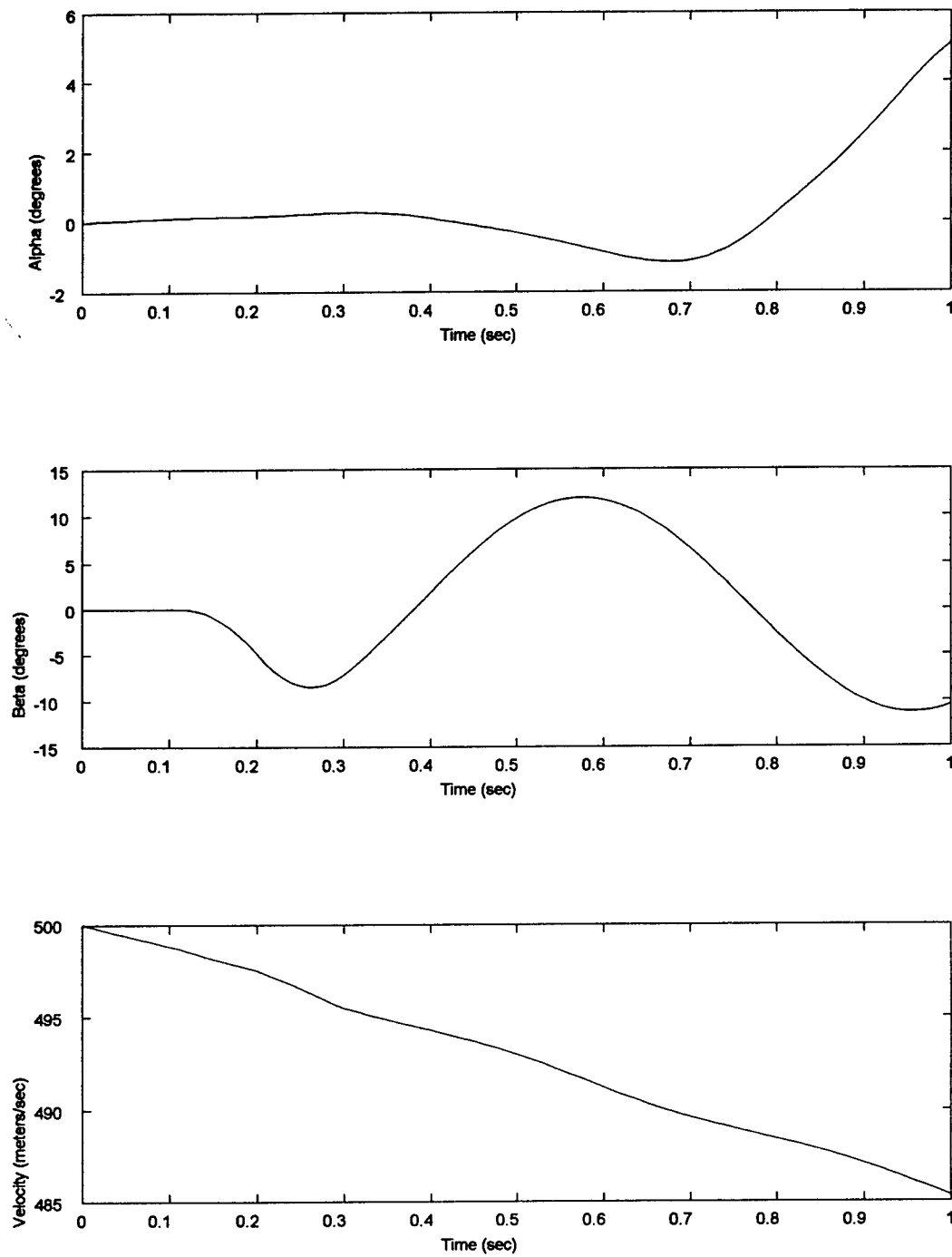


Fig 2.14: Angle-of-attack α (top), side-slip β (middle), and velocity (bottom) vs. time for doublet yaw fin deflection

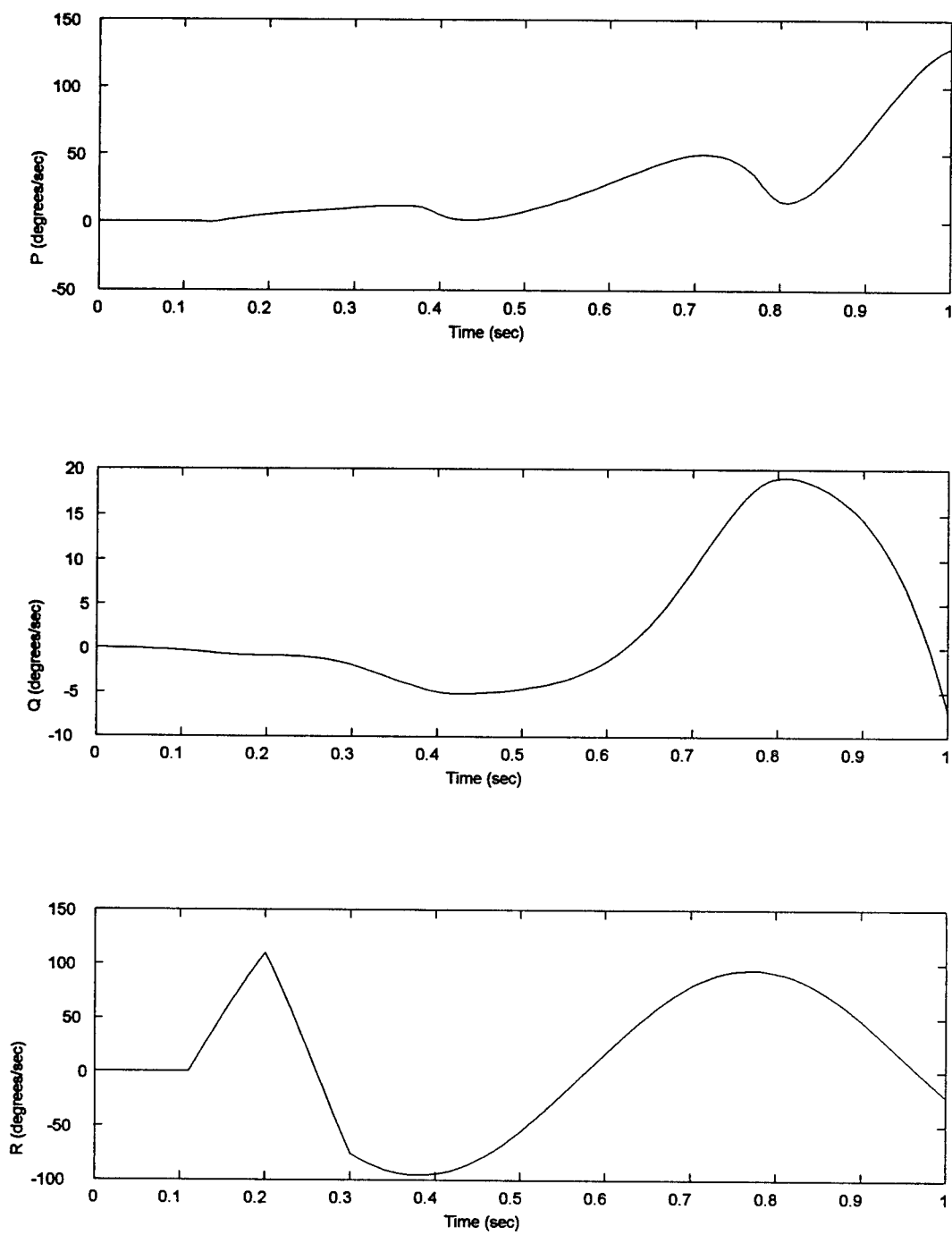


Fig 2.15: Roll rate P (top), pitch rate Q (middle), and yaw rate R (bottom) vs. time for doublet yaw fin deflection

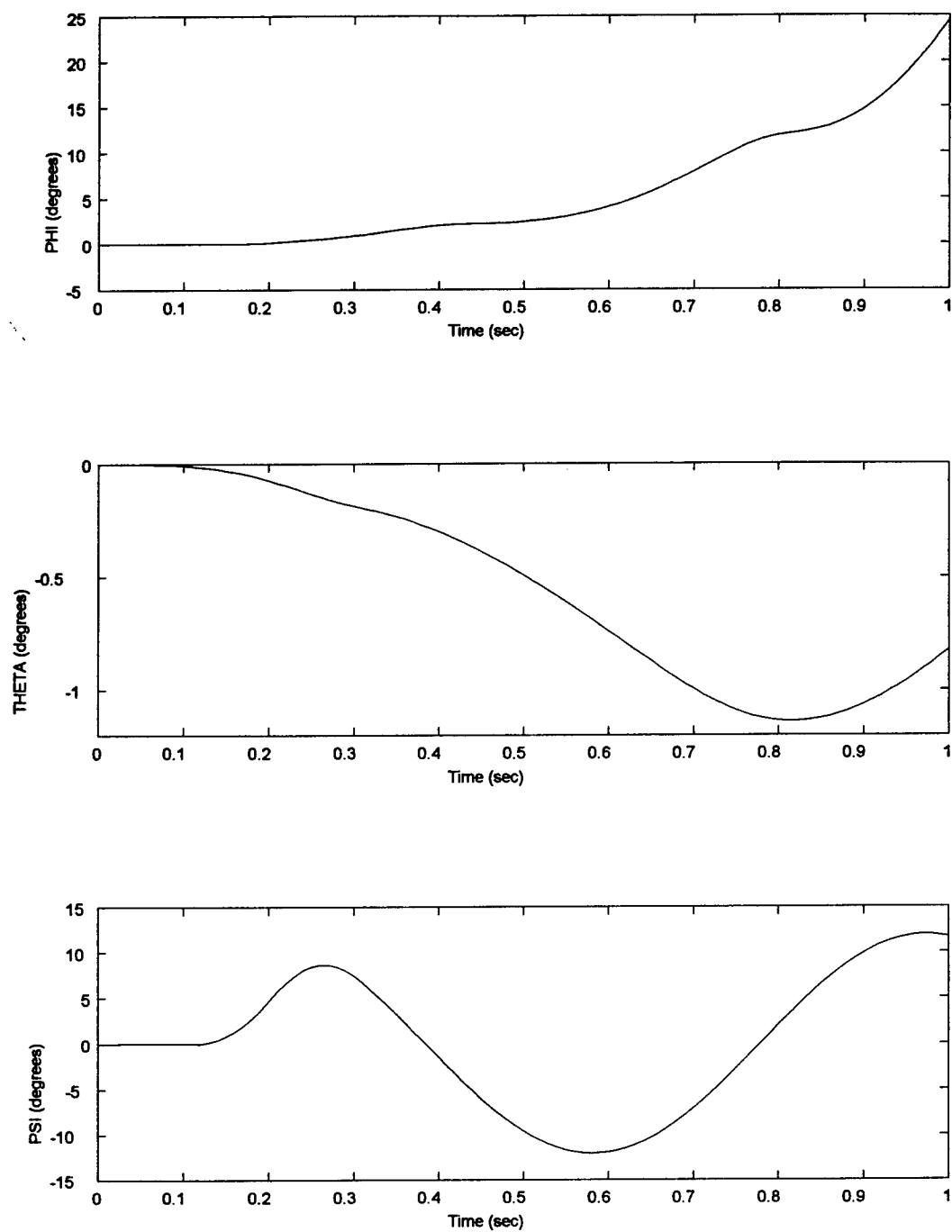


Fig 2.16: Euler angles ϕ (top), θ (middle), and ψ (bottom) vs. time for doublet pitch yaw deflection

It is also of interest to have the capability to obtain linearized models of a missile's nonlinear aerodynamics about user-selected operating points. These models serve as the starting point for the linear design phase of the overall gain scheduling design process. For the air-to-air missile model considered here, linearized models can be obtained in the *Matlab*TM workspace using the *jacob* command. The on-line help message describing proper syntax for this command are shown in Figure 2.17.

```

» help jacob

This function calculates state-space coefficient matrices
for a linearization of the six degree-of-freedom air-to-air
missile model at a user-selected operating point (x,u) where

x = [alpha beta vt phi theta psi P Q R h]'
u = [delp delr dely]'

The linearized model corresponds to state and input vectors
specified above and an output vector containing total accelerations
(including gravitational effects) along body-fixed x, y, and z-axes.

Usage: [a,b,c,d] = jacob(x,u)

Inputs:    x,u : state and input vectors specifying the operating
              point values of the variables indicated above

Outputs: a,b,c,d : coefficient matrices of the linearized state equation
              (a is 10 x 10, b is 10 x 3, c is 3 x 10, d is 3 x 3)

Steven Andrew
12-16-97 Version 1.2

»

```

Fig 2.17: *Matlab* on-line help message for the *jacob* command

As an example of *jacob* command usage, consider the operating point for which *x* corresponds to the initial condition values listed in Table 2.1 and *u* corresponds to zero control surface deflections. The dimensions of the coefficient matrices returned by the *jacob* command preclude the display of these matrices. However, by dividing the ten state variables into longitudinal and lateral components according to

$$x_{long} = \begin{bmatrix} \alpha \\ Q \\ V_t \\ \theta \\ h \end{bmatrix}, \quad x_{lat} = \begin{bmatrix} \beta \\ P \\ R \\ \phi \\ \psi \end{bmatrix}, \quad x = \begin{bmatrix} x_{long} \\ x_{lat} \end{bmatrix}$$

the A and B matrices have the form

$$A = \begin{bmatrix} A_{long} & A_{lat-long} \\ 0 & A_{lat} \end{bmatrix}, \quad B = \begin{bmatrix} B_{long} & 0 \\ 0 & B_{lat} \end{bmatrix}$$

where, for example,

$$A_{long} = \begin{bmatrix} -0.2402 & 1 & 0 & 0 & 0 \\ -62.31 & -0.2836 & 0 & 0 & 0 \\ 9.256 & 0 & -0.0432 & -9.8 & 0.0015 \\ 0 & 1 & 0 & 0 & 0 \\ -500 & 0 & 0 & 500 & 0 \end{bmatrix}, \quad B_{long} = \begin{bmatrix} -0.2006 \\ -140.7 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

In addition, transfer functions corresponding to selected input-output pairs can be calculated. Denoting the effective roll, pitch, and yaw control surface deflections by δ_r, δ_p , and δ_y , respectively, and the roll, pitch and yaw body rates by P, Q , and R , respectively, the following is a representative sample:

$$\begin{aligned} \frac{P}{\delta_r}(s) &= \frac{1.283}{s+21.4} \\ \frac{Q}{\delta_p}(s) &= \frac{-140.7 s(s-0.00379)(s+0.0525)(s+0.146)}{(s-0.0188)(s+0.031 \pm j0.0242)(s+0.262 \pm j7.89)} \\ \frac{R}{\delta_y}(s) &= \frac{-140.7}{s+0.284} \end{aligned}$$

3. Stability Analysis of Gain Scheduled Missile Autopilots

The gain scheduling methodology described earlier assumes that a family of linearized plants has been computed about a family of constant operating points or equilibria. The objective is then to construct a nonlinear controller with the property that the resulting nonlinear closed-loop system has a prescribed family of linearizations about its constant operating point family corresponding to constant exogenous signals. Once this has been achieved, the nonlinear closed-loop system will exhibit bounded behavior for non-constant but sufficiently slowly varying exogenous signals.

This underlying assumption is routinely violated when gain scheduling is employed in the design of longitudinal missile autopilots. Specifically, the starting point for the design is the short-period aerodynamics described by

$$\begin{aligned}\dot{\alpha} &= \frac{1}{V_m} \left[\frac{\bar{q}S}{m} [C_N \cos(\alpha) - C_A \sin(\alpha)] + g \cos(\theta - \alpha) \right] + Q \\ \dot{Q} &= \frac{\bar{q}Sd}{I_y} C_M\end{aligned}\quad (3.1)$$

where α is angle-of-attack, Q is pitch rate, θ is pitch attitude, V_m is missile speed, \bar{q} is dynamic pressure, and C_A , C_N , C_M are axial force, normal force, and pitching moment aerodynamic coefficients, respectively. Not explicitly shown in these equations is the dependence of the aerodynamic coefficients on the elevator deflection, δ . A control surface actuator that produces an elevator deflection in response to a commanded value δ_c is often modeled by a linear system of the form

$$\frac{d}{dt} \begin{bmatrix} \delta \\ \dot{\delta} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\omega_a^2 & -2\zeta_a \omega_a \end{bmatrix} \begin{bmatrix} \delta \\ \dot{\delta} \end{bmatrix} + \begin{bmatrix} 0 \\ \omega_a^2 \end{bmatrix} \delta_c \quad (3.2)$$

Normal acceleration is usually the variable to be controlled and is given by

$$\eta_z = \frac{\bar{q}S}{m} C_N + g \cos(\theta) \quad (3.3)$$

The gain scheduling methodology applied to the short period aerodynamics and the actuator dynamics yields a nonlinear controller of the general form

$$\begin{aligned}\dot{x}_C &= a(x_C, \eta_c, y_m, \Theta) \\ \delta_c &= c(x_C, \eta_c, y_m, \Theta)\end{aligned}\quad (3.4)$$

where x_C is the controller state, η_c is commanded normal acceleration, y_m is a vector of measurements available to the controller, and Θ denotes the vector of scheduling variables. The closed-loop system formed by the interconnection of (3.1), (3.2), and (3.4) can be cast in the general form

$$\Sigma_a: \quad \begin{aligned}\dot{x}_a &= f_a(x_a, u_a) \\ y_a &= h_a(x_a)\end{aligned}\quad (3.5)$$

where

$$x_a = \begin{bmatrix} \alpha \\ Q \\ \delta \\ \dot{\delta} \\ x_C \end{bmatrix} \quad u_a = \begin{bmatrix} V_m \\ h \\ \theta \\ \eta_c \end{bmatrix} \quad y_a = \begin{bmatrix} \alpha \\ Q \\ \delta \end{bmatrix}. \quad (3.6)$$

Although not shown explicitly, the dependence on altitude, h , comes from several sources including altitude dependence of air density and hence dynamic pressure, possible dependence of the aerodynamic coefficients on altitude and mach number, and altitude may be one of the scheduling variables.

As a result of the gain scheduling design process, (3.5) should possess a family of exponentially stable constant operating points that can be parameterized by constant values of the input u_a . Then, if u_a is a bounded and sufficiently slowly varying *exogenous* input signal, the resulting state and output

trajectories will remain bounded. However, in this situation the input u_a to (3.5) is not composed of exogenous signals but instead consists of quantities that also govern the longitudinal aerodynamic behavior. Specifically, V_m , θ , and h evolve according to

$$\begin{aligned}\dot{V}_m &= \frac{\bar{q}S}{m} [C_A \cos(\alpha) + C_N \sin(\alpha)] - g \sin(\theta - \alpha) \\ \dot{\theta} &= Q \\ \dot{h} &= V_m \sin(\theta - \alpha)\end{aligned}\tag{3.7}$$

and assuming constant normal acceleration commands,

$$\dot{\eta}_c = 0.\tag{3.8}$$

The dynamics (3.7), (3.8) can also be represented in the general form

$$\Sigma_b: \begin{aligned}\dot{x}_b &= f_b(x_b, u_b) \\ y_b &= h_b(x_b)\end{aligned}\tag{3.9}$$

where

$$x_b = \begin{bmatrix} V_m \\ Q \\ h \\ \eta_c \end{bmatrix} \quad u_b = \begin{bmatrix} \alpha \\ Q \\ \delta \end{bmatrix} \quad y_b = \begin{bmatrix} V_m \\ h \\ \theta \\ \eta_c \end{bmatrix}.\tag{3.10}$$

The definitions (3.6) and (3.10) imply the following interconnection of the subsystems Σ_a and Σ_b :

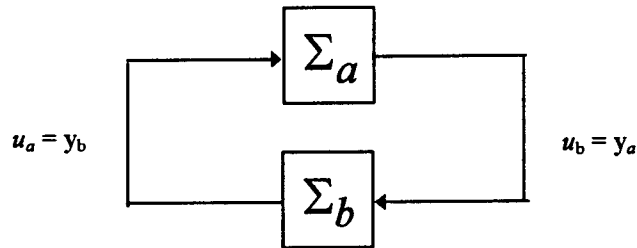


Fig. 3.1: Feedback interconnection

The existing stability analysis that establishes the boundedness property of x_a and y_a for a slowly-varying *exogenous* input u_a cannot be directly applied in this situation. A simple example illustrates that the coupling from Σ_a to Σ_b can have a destabilizing effect. Suppose the subsystem Σ_a is the bounded-input, bounded-output stable linear system given by

$$\begin{aligned}\dot{x}_a &= -x_a + u_a \\ y_a &= x_a\end{aligned}$$

and the subsystem Σ_b is the marginally stable linear system

$$\begin{aligned}\dot{x}_b &= \varepsilon u_b \\ y_b &= x_b\end{aligned}$$

where the parameter ε characterizes the coupling from Σ_a to Σ_b . When $\varepsilon = 0$, there is no such coupling and Σ_b can be viewed as an exosystem that generates constant input signals for Σ_a . Consequently, the corresponding response $y_a = x_a$ will be bounded. For $\varepsilon \neq 0$, the interconnected system is exponentially stable for ε negative and unstable for ε positive, regardless of how small.

In order to develop a stability analysis for the general class of systems at hand in which the previously mentioned slowly varying results can be incorporated, a modification to the basic set-up depicted in Figure 3.1 is first described. Then, standard notions of input-output stability will be employed in order to invoke a version of the well-known small-gain theorem [1] [3] [8] for this modified system.

First, it will be assumed that the subsystem Σ_a of the form (3.5) has a family of exponentially stable operating points that can be parameterized by constant values of the input u_a . Specifically, it will be assumed that there is a continuously differentiable map $x_a^\circ(\alpha)$ defined on a compact set Γ containing the origin that satisfies

$$f_a(x_a^\circ(\alpha), \alpha) = 0 \text{ for all } \alpha \in \Gamma.$$

Further, it will be assumed that there is a positive constant λ such that the eigenvalues of

$$\frac{\partial f_a}{\partial x_a}(x_a^\circ(\alpha), \alpha)$$

have real part no greater than $-\lambda$ for all $\alpha \in \Gamma$. Under these assumptions, it has been shown [2], [4], [7] that, for some positive constant r_a , there exists a continuously differentiable map $V(z, \alpha)$ defined on $B_{r_a} \times \Gamma$ that satisfies

$$\begin{aligned}c_1 \|z\|^2 &\leq V(z, \alpha) \leq c_2 \|z\|^2 \\ \frac{\partial V}{\partial z}(z, \alpha) f_a(z + x_a^\circ(\alpha), \alpha) &\leq -c_3 \|z\|^2 \\ \left\| \frac{\partial V}{\partial z}(z, \alpha) \right\| &\leq c_4 \|z\| \\ \left\| \frac{\partial V}{\partial \alpha}(z, \alpha) \right\| &\leq c_5 \|z\|^2\end{aligned}$$

for all $z \in B_{r_a}, \alpha \in \Gamma$ for positive constants c_1 through c_5 .

Related to the subsystem Σ_a is the following subsystem to which the previously mentioned slowly varying analysis will be applied.

$$\begin{aligned}\tilde{\Sigma}_a: \quad \dot{z}_a &= \tilde{f}_a(z_a, \alpha_a, v_a) \\ \dot{\alpha}_a &= v_a \\ y_a &= \tilde{h}_a(z_a, \alpha_a)\end{aligned}$$

where $z_a = x_a - x_a^\circ(\alpha_a)$,

$$\tilde{f}_a(z_a, \alpha_a, v_a) = f_a(z_a + x_a^\circ(\alpha_a), \alpha_a) - \frac{\partial x_a^\circ}{\partial \alpha} v_a$$

and

$$\tilde{h}_a(z_a, \alpha_a) = h_a(z_a + x_a^\circ(\alpha_a)).$$

Note that, in addition to a change of state coordinates, $\tilde{\Sigma}_a$ results from cascading integrators on the input channels of Σ_a . Consequently, if the outputs of Σ_b are differentiated to yield the subsystem

$$\tilde{\Sigma}_b: \begin{aligned} \dot{x}_b &= f_b(x_b, u_b) \\ w_b &= \tilde{h}_b(x_b, u_b) \end{aligned}$$

where $w_b = \dot{y}_b$ and $\tilde{h}_b(x_b, u_b) = [\partial h_b / \partial x_b](x_b) f_b(x_b, u_b)$, the feedback interconnection of $\tilde{\Sigma}_a$ and $\tilde{\Sigma}_b$ is equivalent as far as the output y_a is concerned provided the newly introduced integrators are initialized properly. Next, the following notion of input-output stability is of interest.

Definition. A system Σ with input u and output y is said to be small-signal, finite-gain L_∞ -stable if there is a positive constant r and nonnegative constants γ and β such that

$$\|y\|_\infty \leq \gamma \|u\|_\infty + \beta$$

for all $u \in L_\infty$ with $\|u\|_\infty \leq r$.

□

The well-known small-gain theorem can be specialized to the feedback interconnection of $\tilde{\Sigma}_a$ and $\tilde{\Sigma}_b$ to yield the following result.

Theorem. Suppose the subsystems $\tilde{\Sigma}_a$ and $\tilde{\Sigma}_b$ are each small-signal, finite-gain L_∞ -stable. That is, there is a positive constant r and nonnegative constants $\gamma_a, \beta_a, \gamma_b$, and β_b such that

$$\|y_a\|_\infty \leq \gamma_a \|v_a\|_\infty + \beta_a \text{ for all } v_a \in L_\infty \text{ with } \|v_a\|_\infty \leq r$$

and

$$\|w_b\|_\infty \leq \gamma_b \|u_b\|_\infty + \beta_b \text{ for all } u_b \in L_\infty \text{ with } \|u_b\|_\infty \leq r.$$

Then, if $\gamma_a \gamma_b < 1$ there holds

$$\|y_a\|_\infty \leq \frac{\gamma_a \beta_b + \beta_a}{1 - \gamma_a \gamma_b}.$$

□ □ □

Next, the desired stability property of the subsystem $\tilde{\Sigma}_a$ is established.

Theorem. Under the above assumptions on the subsystem $\tilde{\Sigma}_a$, if $\|z_a(0)\| < (c_1/c_2)^{1/2} r_a$ and $\alpha_a(0)$ and $v_a(t)$ are such that $\alpha_a(t) \in \Gamma$ for all $t \geq 0$, then $\tilde{\Sigma}_a$ is small-signal, finite-gain L_∞ -stable.

Proof. The time-derivative of $V(z, \alpha)$, introduced above, along trajectories of $\tilde{\Sigma}_a$ satisfies

$$\begin{aligned}\dot{V}(z_a, \alpha_a) &= \frac{\partial V}{\partial z}(z_a, \alpha_a) \tilde{f}_a(z_a, \alpha_a, v_a) + \frac{\partial V}{\partial \alpha}(z_a, \alpha_a) v_a \\ &= \frac{\partial V}{\partial z}(z_a, \alpha_a) \left[f_a(z_a + x_a^\circ(\alpha_a), \alpha_a) - \frac{\partial x_a^\circ}{\partial \alpha}(\alpha_a) v_a \right] + \frac{\partial V}{\partial \alpha}(z_a, \alpha_a) v_a \\ &\leq -c_3 \|z_a\|^2 + \left[\frac{\partial V}{\partial \alpha}(z_a, \alpha_a) - \frac{\partial V}{\partial z}(z_a, \alpha_a) \frac{\partial x_a^\circ}{\partial \alpha}(\alpha_a) \right] v_a \\ &\leq -c_3 \|z_a\|^2 + \left[c_5 \|z_a\|^2 + c_4 K_a \|z_a\| \right] \|v_a\| \quad \forall z_a \in B_{r_a}, \alpha_a \in \Gamma\end{aligned}$$

where, from previous assumptions,

$$K_a = \max_{\alpha \in \Gamma} \left\| \frac{\partial x_a^\circ}{\partial \alpha}(\alpha) \right\| < \infty.$$

If, for $\theta \in (0, 1)$,

$$\|v_a\|_\infty \leq (1-\theta) \frac{c_3}{c_5}$$

then

$$\begin{aligned}\dot{V}(z_a, \alpha_a) &\leq -\theta c_3 \|z_a\|^2 + c_4 K_a \|z_a\| \|v_a\| \\ &\leq -\theta \frac{c_3}{c_2} V(z_a, \alpha_a) + \frac{c_4}{\sqrt{c_1}} K_a \|v_a\| \sqrt{V(z_a, \alpha_a)}\end{aligned}$$

Letting $W(t) = \sqrt{V(z_a(t), \alpha_a(t))}$ it can be shown that

$$\begin{aligned}D^+ W(t) &= \limsup_{h \rightarrow 0^+} \frac{W(t+h) - W(t)}{h} \\ &\leq -\frac{\theta}{2} \frac{c_3}{c_2} W(t) + \frac{c_4 K_a}{2\sqrt{c_1}} \|v_a(t)\|\end{aligned}$$

from which the comparison lemma [3, Lemma 2.5] yields

$$W(t) \leq e^{-\frac{\theta c_3}{2 c_2} t} W(0) + \frac{c_4 K_a}{2\sqrt{c_1}} \int_0^t e^{-\frac{\theta c_3}{2 c_2} (t-\tau)} \|v_a(\tau)\| d\tau$$

which in turn gives

$$\|z_a(t)\| \leq \sqrt{\frac{c_2}{c_1}} e^{-\frac{\theta c_3}{2 c_2} t} \|z_a(0)\| + \frac{c_4 K_a}{2 c_1} \int_0^t e^{-\frac{\theta c_3}{2 c_2} (t-\tau)} \|v_a(\tau)\| d\tau.$$

The bounds $\|z_a(0)\| < \sqrt{\frac{c_1}{c_2}} r_a$ and $\|v_a\|_\infty < \frac{c_1 c_3 \theta}{c_2 c_4 K_a} r_a$ ensure that $\|z_a(t)\| < r_a, \forall t \geq 0$ so that the

preceding analysis is valid from which

$$\|z_a\|_\infty \leq \frac{c_2 c_4 K_a}{c_1 c_3 \theta} \|v_a\|_\infty + \sqrt{\frac{c_2}{c_1}} \|z_a(0)\|.$$

Then, since $x_a = z_a + x_a^\circ(\alpha_a)$ and by hypothesis $\alpha_a(t) \in \Gamma$ for all $t \geq 0$,

$$\|x_a\|_\infty \leq \|z_a\|_\infty + \max_{\alpha \in \Gamma} \|x_a^\circ(\alpha)\|.$$

Finally, under the assumption that the map $h_a(x_a)$ is continuously differentiable, there is a finite constant L_a such that

$$\|h_a(x_a)\| \leq L_a \|x_a\| \quad \forall x_a \text{ with } \|x_a\| \leq r_a + \max_{\alpha \in \Gamma} \|x_a^\circ(\alpha)\|.$$

It follows that for

$$\|v_a\|_\infty \leq r = \min \left\{ (1-\theta) \frac{c_3}{c_5}, \frac{c_1 c_3 \theta}{c_2 c_4 K_a} r_a \right\}$$

there holds

$$\|y_a\|_\infty \leq \gamma_a \|v_a\|_\infty + \beta_a$$

with

$$\gamma_a = L_a \frac{c_2 c_4 K_a}{c_1 c_3 \theta}, \quad \beta_a = L_a \left(\sqrt{\frac{c_2}{c_1}} \|z_a(0)\| + \max_{\alpha \in \Gamma} \|x_a^\circ(\alpha)\| \right).$$

□ □ □

Returning to the longitudinal missile system, the subsystem $\tilde{\Sigma}_a$ corresponding to (3.1), (3.2), (3.3), and (3.4) will possess the input-output stability property established in the preceding theorem by virtue of the gain scheduling design process. Input-output stability of the subsystem $\tilde{\Sigma}_b$ corresponding to (3.7) and (3.8) will now be addressed. First, consider the energy function

$$E(V_m, h) = \frac{1}{2} m V_m^2 + mgh.$$

The time derivative of E along trajectories of (3.7) is given by

$$\begin{aligned} \dot{E}(V_m, h) &= m V_m \dot{V}_m + mg \dot{h} \\ &= -V_m F_D \end{aligned}$$

where $F_D = \bar{q} S C_D$ is the drag force with C_D the drag coefficient related to axial and normal force coefficients via

$$C_D = -[C_A \cos(\alpha) + C_N \sin(\alpha)].$$

Since the drag force is always positive for a nonzero velocity,

$$E(V_m(t), h(t)) < E(V_m(0), h(0)), \quad t \geq 0$$

for nonzero initial velocity and altitude. It can be concluded that $V_m(t)$ and $h(t)$ are bounded for $t \geq 0$ and so the same is true for dynamic pressure. This indicates $\dot{h}(t)$ is bounded for $t \geq 0$ and, under

the realistic assumption that the drag coefficient above can be bounded uniformly in angle-of-attack, pitch rate, and elevator deflection, boundedness of dynamic pressure leads to boundedness of $\dot{V}_m(t)$ for $t \geq 0$. This gives, using (3.7) and (3.8),

$$\begin{bmatrix} \dot{V}_m(t) \\ \dot{\theta}(t) \\ \dot{h}(t) \\ \dot{\eta}_c(t) \end{bmatrix} \leq \|Q(t)\| + \beta_b \leq \begin{bmatrix} \alpha(t) \\ Q(t) \\ \delta(t) \end{bmatrix} + \beta_b, \quad t \geq 0$$

for finite β_b . Finally, with $w_b = [\dot{V}_m \quad \dot{\theta} \quad \dot{h} \quad \dot{\eta}_c]^T$ and u_b as in (3.10), there holds

$$\|w_b\|_\infty \leq \gamma_b \|u_b\|_\infty + \beta_b$$

with β_b as in the previous inequality and $\gamma_b = 1$.

The preceding analysis provides an explanation for the performance exhibited by longitudinal gain scheduled missile autopilots designed on the basis of linearizations about non-equilibrium conditions. The complete longitudinal dynamics can be cast as a feedback interconnection of two subsystems ($\tilde{\Sigma}_a$ and $\tilde{\Sigma}_b$). Input-output stability of $\tilde{\Sigma}_a$ has been argued using analysis techniques for systems with sufficiently slowly varying input signals. Input-output stability of $\tilde{\Sigma}_b$ has been established in an ad-hoc manner using physical insight. Bounded behavior of the overall system follows from a version of the small-gain theorem. This analysis is not intended to provide quantitative conditions for bounded behavior, but rather to validate in a general sense the potentially hazardous approach of designing controllers on the basis of linearized models calculated about non-equilibrium conditions.

4. Concluding Remarks

The computational tools described in Section 2 constitute preliminary efforts to develop a general-purpose environment for the analysis and design of gain scheduled missile autopilots. A long-term objective is a systematic procedure whereby, starting with tabular aerodynamic data for any flight vehicle, nonlinear six degree-of-freedom simulations and trimming and linearization routines can quickly and easily be constructed. At the present time, such capabilities are restricted to a specific air-to-air missile model, but the experience gained from this exercise will serve as a springboard for the development of a completely general framework in the future.

The stability analysis of Section 3 lends justification to the potentially hazardous, yet commonly practiced technique of designing linear autopilots based on linearizations of a missile's nonlinear aerodynamics about non-equilibrium conditions. The analysis indicates that when a collection of linear autopilots designed in this manner is used to construct a gain scheduled autopilot, bounded system performance can be expected, thus confirming nonlinear simulation results and actual behavior in the air.

5. References

- [1] C. A. Desoer and M. Vidyasagar, *Feedback Systems: Input-Output Properties*, Academic Press, New York, 1975.
- [2] M. Kelemen, "A Stability Property," *IEEE Transactions on Automatic Control*, Vol. 31, No. 8, pp. 766-768, 1986.
- [3] H. K. Khalil, *Nonlinear Systems*, Second Edition, Prentice-Hall, New Jersey, 1996.
- [4] H. K. Khalil and P. V. Kokotovic, "On Stability Properties of Nonlinear Systems with Slowly Varying Inputs, *IEEE Transactions on Automatic Control*, Vol. 36, No. 2, p. 229, 1991.
- [5] D. A. Lawrence, "Analysis and Design of Gain Scheduled Missile Autopilots," AFOSR Summer Faculty Research Program Final Report, Wright Laboratory, Eglin AFB, August 1996.
- [6] D. A. Lawrence and W. J. Rugh, "Gain Scheduling Dynamic Linear Controllers for a Nonlinear Plant," *Automatica*, Vol. 31, No. 3, pp. 381-390, 1995.
- [7] D. A. Lawrence and W. J. Rugh, "On a Stability Theorem for Nonlinear Systems with Slowly Varying Inputs, *IEEE Transactions on Automatic Control*, Vol. 35, No. 7, pp. 860-864, 1990.
- [8] M. Vidyasagar, *Nonlinear Systems Analysis*, Second Edition, Prentice-Hall, New Jersey, 1993.

**DETERMINATION OF 3D DEFORMATIONS, FORCES AND MOMENTS OF
AIRCRAFT TIRES WITH A SYNCHRONIZED OPTICAL AND ANALOG SYSTEM**

Junghsen Lieh
Associate Professor
Mechanical & Materials Engineering

Wright State University
3640 Colonel Glenn Highway
Dayton, Ohio 45435

Final Report for:
Summer Research Extension Program
Wright Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Wright Laboratory

December 1997

DETERMINATION OF 3D DEFORMATIONS, FORCES AND MOMENTS OF AIRCRAFT TIRES WITH A SYNCHRONIZED OPTICAL AND ANALOG SYSTEM

Junghsen Lieh
Associate Professor
Mechanical & Materials Engineering
Wright State University

Abstract

This report describes the aircraft tire properties measured with Wright State's Optotrak vision system. The data of pre-defined points was detected in a 3D global coordinate system using infrared light emitted from the markers. The optical device is synchronized with analog (non-optical) sensors, i.e., force and torque transducers. In the first experiment, the tire (KC-135 and F-16) was mounted on a Tire Force Machine (TFM), where the vertical, lateral and longitudinal (fore-aft) forces were applied. The instrument recorded the tire deformation, force, moment, TFM tolerance and table movement as the table travels. The second experiment involved the use of an 84" dynamometer.

It was observed that the TFM structure sloppy induces an error as high as 100%. Due to the tolerance and misalignment of the testing machine, the real tire deformation measured from global coordinates has to subtract the TFM structure motion. In other words, for each tire, the data measured in global coordinates must be converted into a body coordinate system in order to eliminate the noise. The dynamometer data was contaminated because of vibration and light deflection of the glass; therefore a retest will be needed.

DETERMINATION OF 3D DEFORMATIONS, FORCES AND MOMENTS OF AIRCRAFT TIRES WITH A SYNCHRONIZED OPTICAL AND ANALOG SYSTEM

Junghsen Lieh

Introduction

A CCD is an electric image sensor with light sensitive pixels. Numerous researchers are continuously improving the device and further explore its commercial potentials. A vision sensor is capable of observing geometry and movement of an object. The images can be recorded with passive or active light sources (regular, laser or infrared) reflecting or emitting from the object. A computerized vision sensor contains the following advantages:

- (1) The light source can be detected at a remote distance;
- (2) The sensor has a high spatial resolution (for example 2048 pixels per array) with adjustable fields of view;
- (3) The measurements may be displayed in real time with hundreds of frames per second;
- (4) The system is portable requiring little human labor.

The measured 3D displacements may be converted into velocity and acceleration vectors of the object movement or into a local coordinate system for further analysis. This data can be used for modal analysis (such as frequencies and amplitudes with Fast Fourier Transform). It can also be utilized for visualization and control signals.

It was observed that the TFM structure sloppy could yield an error in tire deformation by nearly 100%. The real tire deformation measured in global coordinates has to subtract the TFM structure motion. To accomplish this, the raw data must be converted into a body coordinate system thus to eliminate the error induced by the TFM structure tolerance and misalignment.

The tire data for the 84" dynamometer was contaminated by the glass vibration and light deflection; therefore a retest will be needed.

Methodology

This section describes the instrument, the procedure of measurements, coordinate transformations, and results.

1. Synchronized Optical and Analog System

The schematic of the measurement system is shown in Figure 1. The system includes

- A 3D-position sensor with the depth of field ranging from 3 to 20 meters. The highest accuracy is 0.05~0.1 mm in x and y directions and 0.1~0.15 mm in z axis.
- An image control with maximum marker and frame sampling rates of 3,000 and 400 Hz. The maximum number of markers to be detected is 256.
- A data acquisition unit synchronizes 16 analog transducers. A power supply and a system conditioner were located between the analog sensors and the data acquisition unit.
- A PC with an interface board and a monitor. The interface board is connected to the system image control unit. Data is displayed in the monitor and stored in the hard drive.
- Multiple-channel IRED strobers are connected to the system image control unit. IRED signals are collected by the 3-lens sensor and processed by the system image control unit.

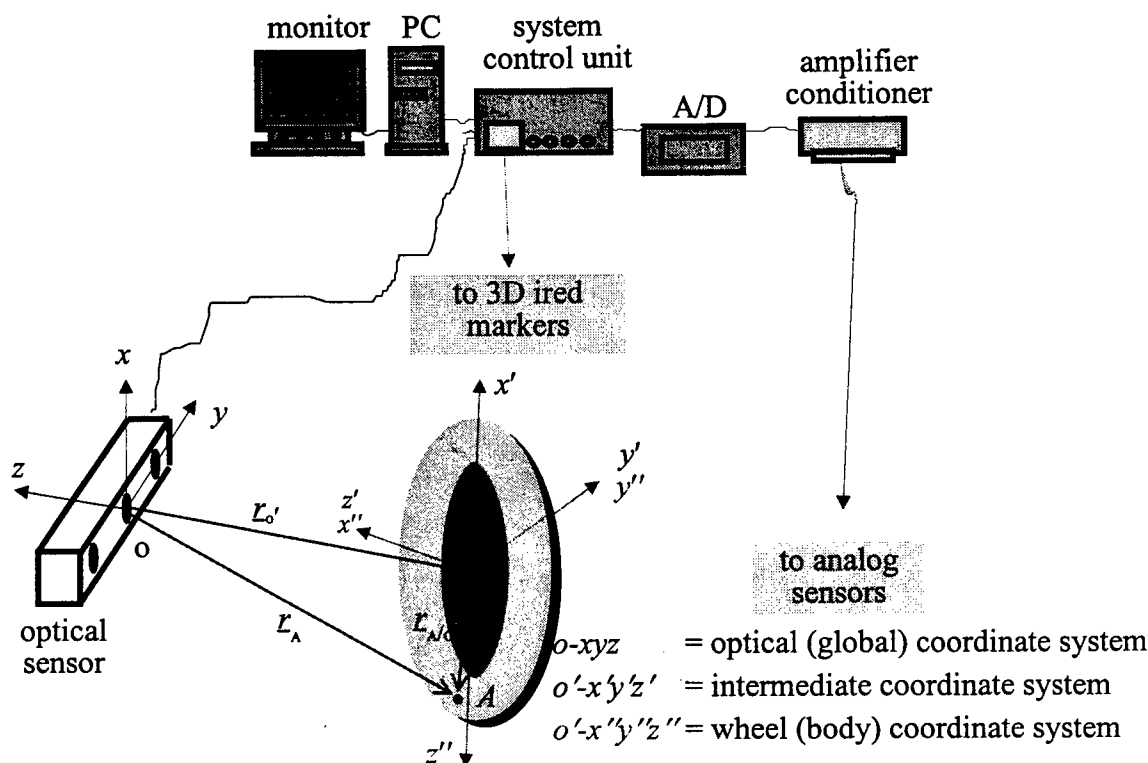


Figure 1. Schematic of the optical and coordinate system for tire measurement.

Prior to testing, active IRED markers were mounted on the tire, wheel, machine frame, and TFM table. The optical device was placed about 3 meters from the tire sidewall. During the lateral testing, the optical system was moved to one end of the TFM table, however, the system was placed at the side of table for the rolling testing. The analog channels for forces, moments, table motion, and contact pressures were connected to a data acquisition unit, which is synchronized with the optical sensor.

2. Data Processing, Coordinate Transformation, and Analysis

The coordinate data represents 3D time histories (i.e., time-domain positions) of the infrared markers mounted on the tire, wheel, machine frame and table. The analog voltage represents forces and moments, which were amplified through calibration factors. To reduce the hard drive memory, all data sets were stored in a binary form, which was later processed into an

ASCII format before any further use. MATLAB programs perform the downstream processing tasks (such as computations and plotting). After the data set has been converted, it was examined and trimmed in order to extract the interested sample points. The final version of the data was analyzed and provided in digital and graphic formats.

To eliminate the tolerance and misalignment in the testing machine thus to obtain true tire deformations, all samples will need corrections through two coordinate transformations: (1) from optical (global) coordinates to intermediate coordinates and (2) from the intermediate coordinate system to wheel coordinates. The transformations involve the computation of Euler angles of the wheel plane for each time step (i.e., each sample). In addition to coordinate transformations, the following tasks may be considered. (1) Process the data into a readable ASCII format to allow the tire data be easily used for scientific and applied research. (2) Construct hysteresis curves. (3) Establish the relations among 3D loading (forces, moments, contact pressures), 3D deformations, coefficients of adhesion, and yaw angles in the wheel coordinate system.

2.1. Coordinate Transformation for Noise Elimination

The maximum TFM frame movement with respect to table directions during the tire test was found as much as follows:

Longitudinal motion:	35 mm (1.4")
Lateral motion:	15 mm (0.6")
Vertical motion:	5 mm (0.2")

The typical time history of lateral and longitudinal wheel center movement during one test is shown in Figures 2 and 3. The TFM misalignment and tolerance obviously induces a significant error if they are included as part of the tire deformation. To eliminate the noise, it will be necessary to normalize the tire data with respect to wheel coordinates through the following coordinate transformations:

- from optical (global) coordinates $o\text{-}xyz$ to an intermediate coordinate system $o'\text{-}x'y'z'$, as shown in Figure 1,
- from the intermediate coordinate system $o'\text{-}x'y'z'$ to wheel coordinates $o''\text{-}x''y''z''$.

Euler angles of the wheel plane for each sample (i.e., each time step) and each measurement are needed, and they were determined through the markers located on the TFM frame. As illustrated in Figure 1, the relative position vector expressed in the $o\text{-}xyz$ (optical or global) coordinate system is

$$\mathbf{r}_{A/o'} = \mathbf{r}_A - \mathbf{r}_{o'} \quad (1)$$

where \mathbf{r}_A and $\mathbf{r}_{o'}$ were measured by the optical system. When expressed in $o'\text{-}x'y'z'$ coordinates, the local position vector becomes

$$\mathbf{r}'_{A/o'} = \mathbf{R}_1 (\mathbf{r}_A - \mathbf{r}_{o'}) \quad (2)$$

where \mathbf{R}_1 is the rotational matrix from $o\text{-}xyz$ to $o'\text{-}x'y'z'$ satisfying the transformation of unit vectors of $o'\text{-}x'y'z'$ and $o\text{-}xyz$ coordinate systems:

$$\mathbf{e}_{o'} = \mathbf{R}_1 \mathbf{e}_o \quad (3)$$

The final data will be converted into the wheel coordinate system ($o''\text{-}x''y''z''$), i.e.,

$$\mathbf{r}''_{A/o'} = \mathbf{R}_2 \mathbf{r}'_{A/o'} \quad (4)$$

where \mathbf{R}_2 is the rotational matrix from $o'\text{-}x'y'z'$ to $o''\text{-}x''y''z''$.

To express the data in minimum memory and processing time, most data sets were trimmed and only the interested points were used. The intention of the task is to provide the data in an easily readable format. In addition, the relation among 3D loading (forces, moments and contact pressures), true tire deformations, coefficients of adhesion, and yaw angles might be established.

Further analysis will be needed beyond the Summer Extension Program because of a large number of data sets (approximately 250 MB). The goal of this project is to provide the basic information for better understanding of the fundamentals of aircraft tire mechanics such that the data may be utilized for more accurate vehicle simulation and design.

The following sections are the results of such analyses in both untransformed and transformed coordinate systems. The result of KC-135 tire is given first, and then followed by the F-16 tire.

2.2. KC-135 Tire on Tire Force Machine (TFM)

The tire (SN 60150607) with a undeformed dimension of 48.6" (dia.) \times 16.8" (width) was inflated to 170 psi. Two test plans were performed, i.e., free rolling and lateral deformation tests (see Figures 4 and 5). In both test plans, the normal load was set to 42k lbs, and a sampling rate was given. 17 markers were used, with 10 on the tire sidewall, 4 on the wheel and TFM frame, and 3 on the moving table. In addition to the optical markers, 7 analog channels recorded the 3D tire forces, moments and table position, i.e., F_x'' , F_y'' , F_z'' , M_x'' , M_y'' , M_z'' , and x_T , respectively.

For the free rolling test, the tire plane was aligned with the fore-aft direction of the table (see Figure 4). Figures 6 and 7 show the sidewall deformation of tire with 4°-yaw angle in untransformed and transformed coordinate systems. It can be observed that the untransformed tire deformation contains a significant error.

For the lateral deformation test plan as shown in Figure 5, the tire was rotated 90°. In other words, the tire sidewalls were in the lateral direction of the table. The lateral force applied in this test plan started from 0 to 8k lbs, then the force was reduced back to 0 lbs. The force was continuously applied from 0 to -8k lbs and finally returned to 0 lbs again. The sidewall deformation expressed in terms of global coordinates and wheel coordinates is shown in Figures 8 and 9, in which an error of over 80% was observed in the untransformed data. These figures

indicate that the true sidewall deformation is much smaller after the structural misalignment and machine tolerance have been eliminated.

2.3. *F-16 Tire on Tire Force Machine (TFM)*

An F-16 Goodyear baseline radial block 50 tire (27.75×8.7R14.5/24PR SN 33381603) was inflated to 300 psi. Different test plans were performed. However, only the vertical and lateral-relaxation tests were analyzed due to the limit in the time and large size of data (see Figure 10). Some results from preliminary analysis for the other tests are also presented.

In the vertical test plan, 5 markers were mounted on the sidewall. The wheel was moved downward until the normal load reached 20k lbs, it was then moved upward to its original position. Figures 11 and 12 show the sidewall profile in the global coordinate and wheel coordinate systems. It is evident that without coordinate transformations, the sidewall profile is not physically explainable. The relation between the normal loading and tire radial deflection is shown in Figure 13, in which a nearly linear Hooke's law is observed, although a hysteresis phenomenon exists.

In the lateral-relaxation test, the normal load was set to 20k lbs. There were 18 markers used, including 5 on the tire sidewall, 4 on the wheel and TFM frame, 3 on the moving table, and 5 on the tire periphery (see Figure 10). The table traveled back and forth, with lateral force starting from 0 to -6k lbs, then returned to 0 lbs. The force was continuously applied from 0 to 6k lbs and finally returned 0 lbs again. The untransformed and transformed sidewall deformations are shown in Figures 14 and 15. Since the untransformed deformation includes the TFM frame movement, therefore, an error of over 100% is observed. As the lateral force exceeded certain value, the tire near the wheel interface will buckle (see Figure 15). The tire periphery bending (relaxation) is shown in Figures 16 and 17, again, a significant error exists if the untransformed data is used.

The preliminary results of other testing are illustrated in Figures 18 and 19, in which Figure 18 is tire sidewall deformation for the free rolling test, and Figure 19 is for the tire contact pressure. Further research work will be needed to complete the data processing and analysis in the future, depending on the funding availability.

2.4. F-16 Tire on 84" Dynamometer

For safety consideration, a glass was placed between the optical system and the dynamometer. When machine was operated, it was found that the glass vibrated with the dynamometer. The light going through the glass has been contaminated. It is concluded that the data can no longer be used, and further retest without glass is recommended.

3. Summary and Conclusions

The use of a synchronized optical and analog device allows 3D images and analog data be collected concurrently. The relation between the tire deformation and force/moment under various operating conditions is critical to the vehicle dynamics simulation and control synthesis. Improper tire data can result in significant errors in the vehicle design, which in turn may lead to safety problems. The testing machine itself (particularly the TFM) was found to move with large displacements along the table longitudinal and lateral directions (i.e., 35 mm and 15 mm, respectively). The TFM frame sloppy itself could have been greater than the true tire deformation, especially for small tires (such as F-16 tires). Therefore, any tolerance or misalignment in the testing machine should be eliminated in order to obtain true tire properties.

It may be possible to redesign the TFM mechanism or at least replace some worn-out parts in order to minimize the sloppy. However, this will certainly increase the cost. In addition, no matter how accurate the new mechanism will be, it still expects some tolerance in the machine. To completely eliminate the tire measurement noise, the simplest yet cheapest approach will be to adopt the method used in this project

The data sets collected during the summer of 1996 include the machine tolerances and misalignments. After coordinate transformations, these error can be eliminated. As is expected, the research work can lead to tremendous improvements in the development of accurate and fundamental tire characteristics. Besides, the data may be utilized for verification of existing finite element simulation codes. The establishment of accurate relationships among tire deformations, forces, moments, contact pressures, yaw angles, and coefficients of adhesion will benefit the Air Force and aircraft industry.

4. References

- Beynon JDE and DR Lamb (1980), Charge-Coupled Devices and Their Applications, McGraw-Hill, New York.
- Freeman H (1990), Machine Vision for Three Dimensional Scenes, Academic Press, Boston.
- Gallespie TD, (1992), "Tires," Fundamentals of Vehicle Dynamics, Chapter 10, Society of Automotive Engineers, Warrendale PA.
- Howes MJ and DV Morgan (1979), Charge-Coupled Devices and Systems, Wiley, New York.
- Jenkins TE (1987), Optical Sensing Techniques and Signal Processing, Prentice-Hall, Englewood Cliffs, NJ.
- Lieh J and Mangal C (1996), "Determination of 3D Real-Time Dynamic Deformation of Vehicle Tires," 15th Tire Society Annual Meeting and Conference, Akron, Ohio.
- Lieh J (1996), "Determination of 3D Deformations, Forces, and Moments of Aircraft Tires with a Synchronized Optical and Analog System," Summer Faculty Program Report No. 34, RDL/AFOSR.
- Lieh J and Zendah S and Medzorian J (1997), "Measurement of 3D Tire Deformation, Force, Moment, and Contact Pressure," SAE International Congress and Exposition, Detroit.
- Lieh J and Zendah S and Medzorian J (1997), "Determination of 3D Aircraft Tire Deformation with Optical Sensors," 16th Tire Society Annual Meeting and Conference, Akron, Ohio.
- Vandoren A (1982), Data Acquisition Systems, Reston Pub., Virginia.

Wong JY (1993), "Mechanics of Pnematic Tires," Theory of Ground Vehicles, Chapter 1, 2nd Edition, Wiley, New York.

Yang D and Illingworth J (1993), "Estimating Rigid 3D Motion by Stereo Fixation of Vertices," British Machine Vision Conference, Vol. I, pp. 45-53.

Acknowledgement

The author would like to express his sincere appreciation to Mr. John Medzorian, Mr. Paul Ulrich and Dr. Arnold Mayer of Air Force Research Laboratory for their financial support through AFOSR.

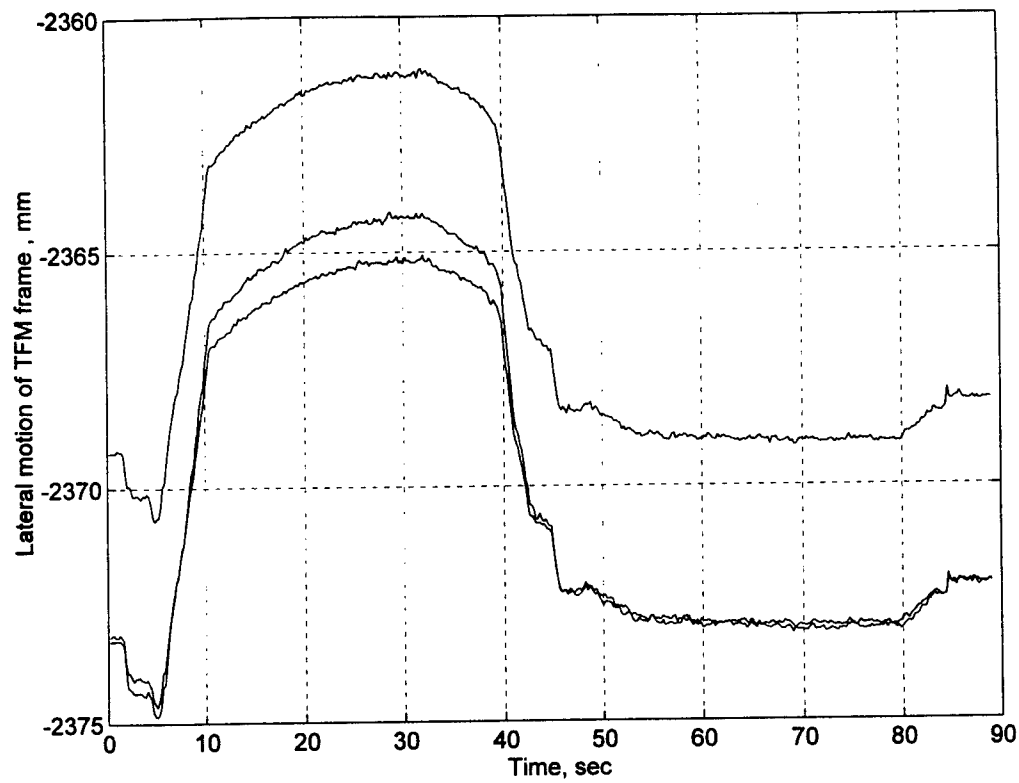


Figure 2. TFM frame lateral movement during test.

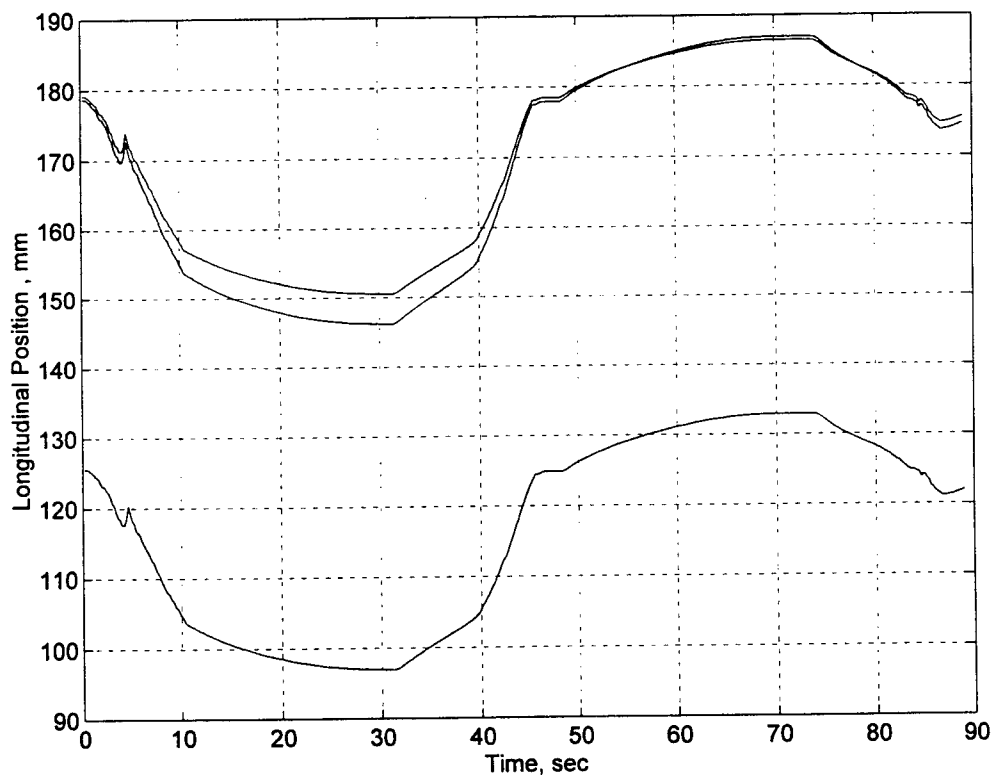


Figure 3. TFM frame longitudinal movement during test.

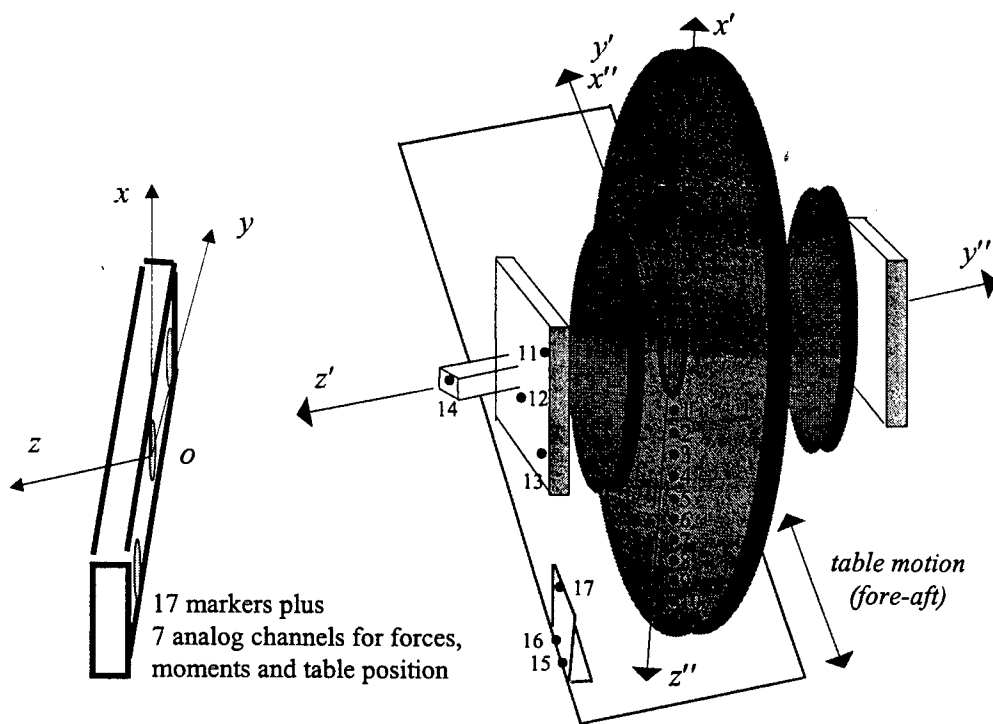


Figure 4. KC-135 tire free rolling testing.

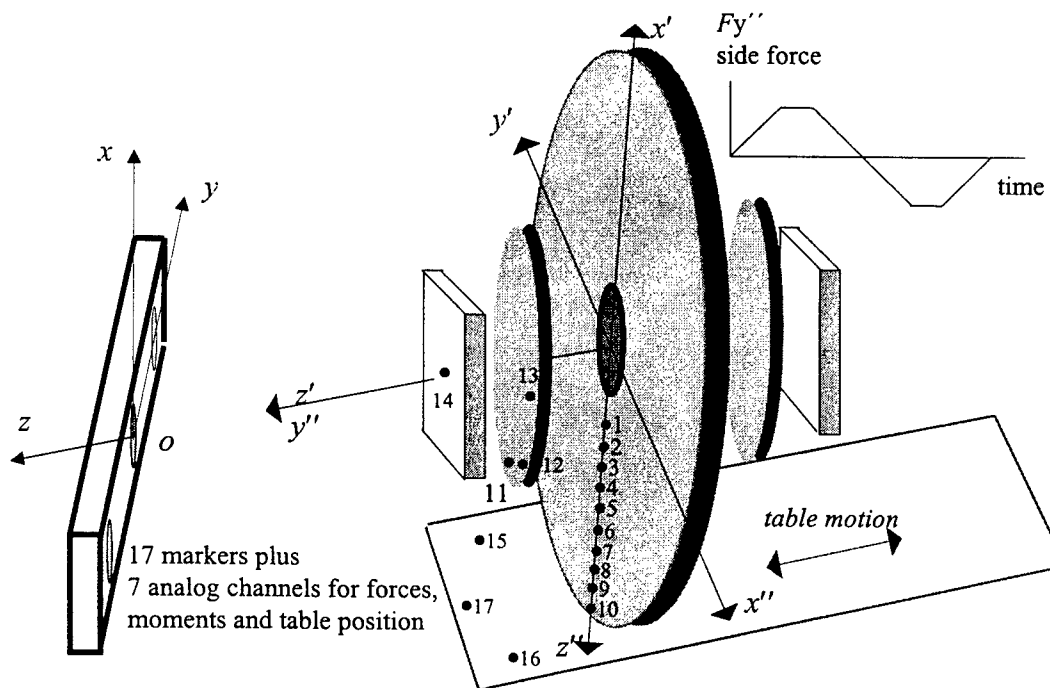


Figure 5. KC-135 tire lateral deformation test.

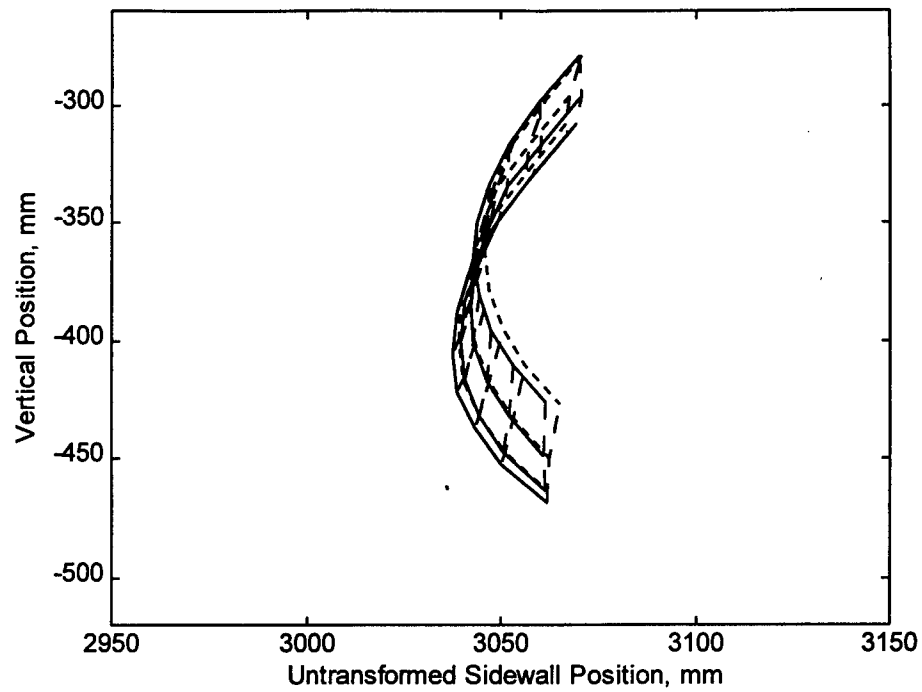


Figure 6. KC-135 tire untransformed sidewall profile under free rolling test.

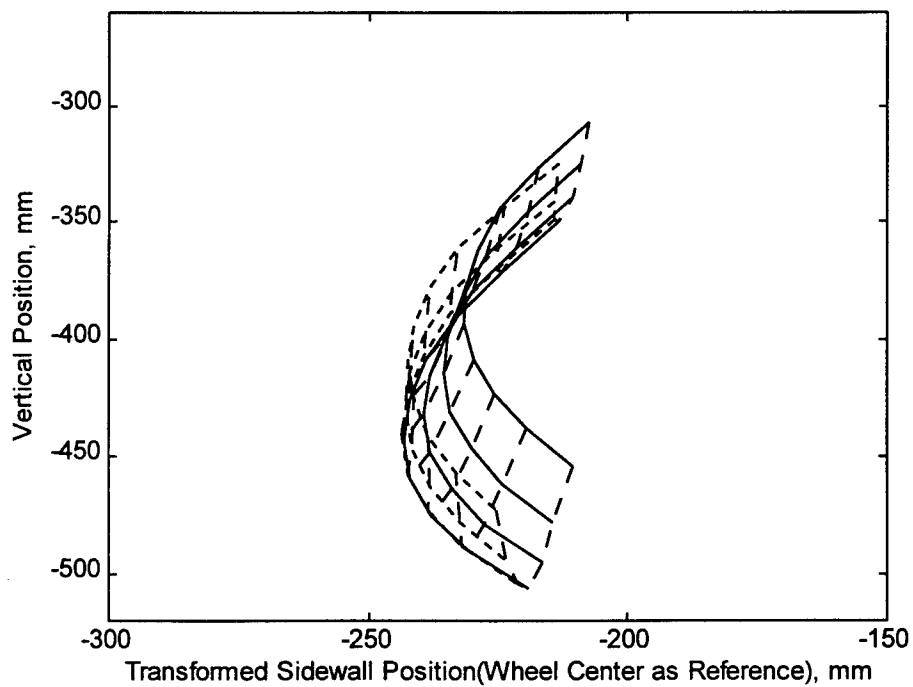


Figure 7. KC-135 tire transformed sidewall profile under free rolling test.

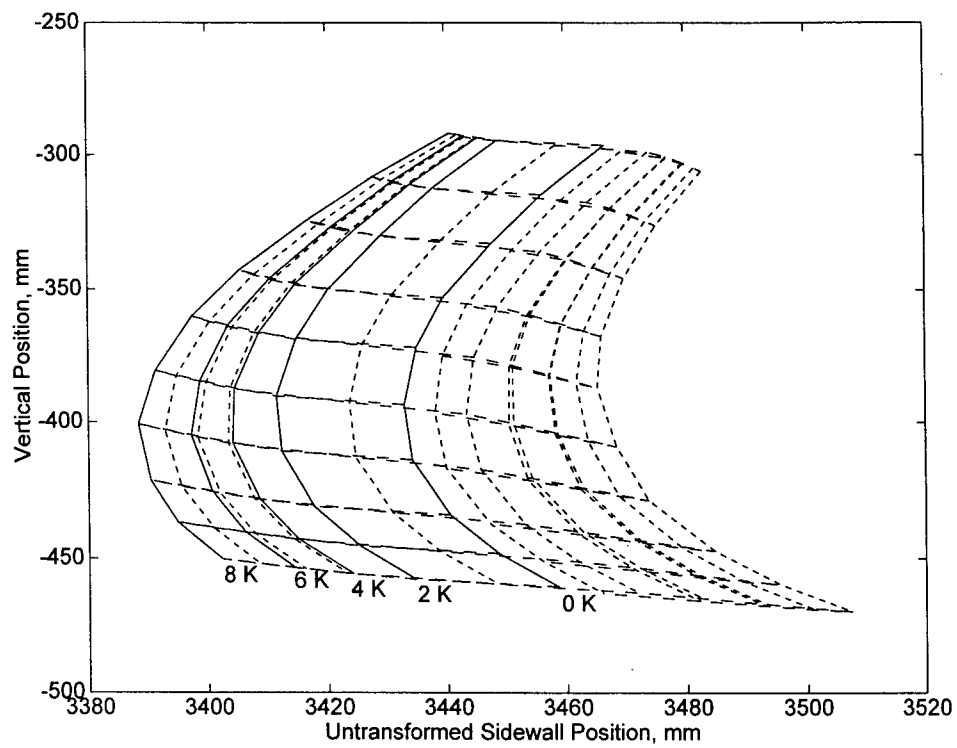


Figure 8. KC-135 tire untransformed sidewall deformation under lateral-relaxation test.

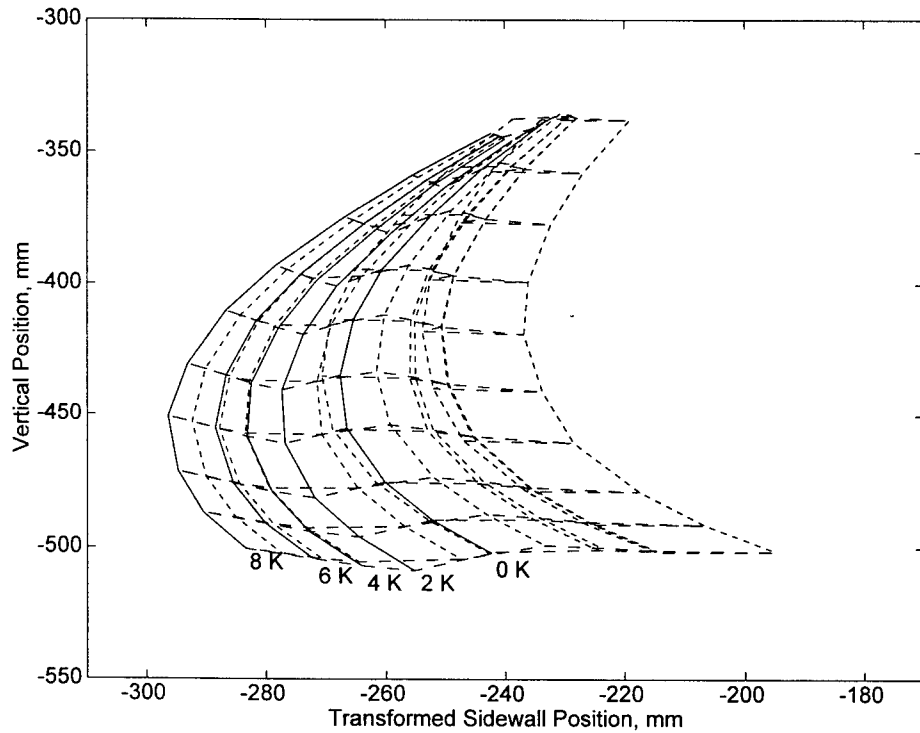


Figure 9. KC-135 tire transformed sidewall deformation under lateral-relaxation test.

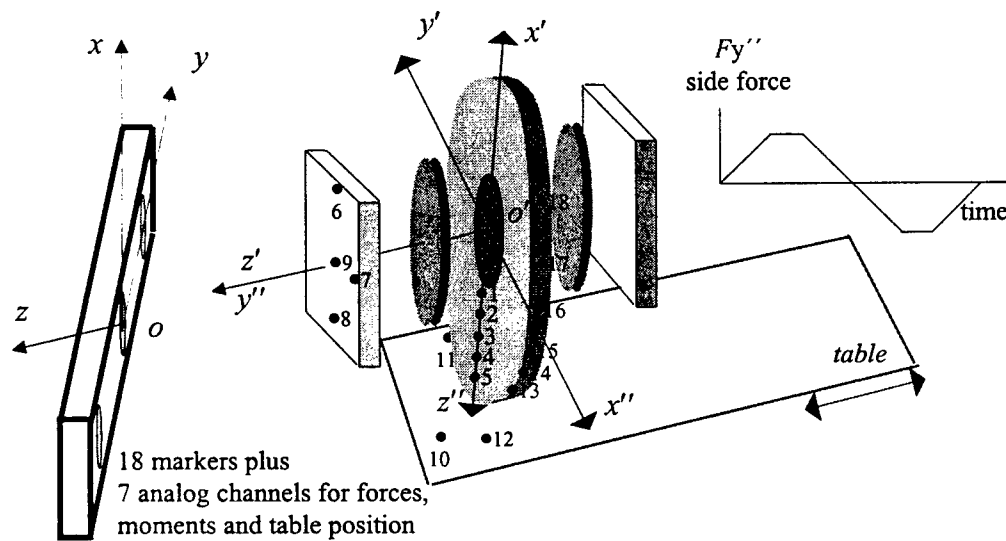


Figure 10. Schematic of F-16 tire under vertical and lateral-relaxation tests.

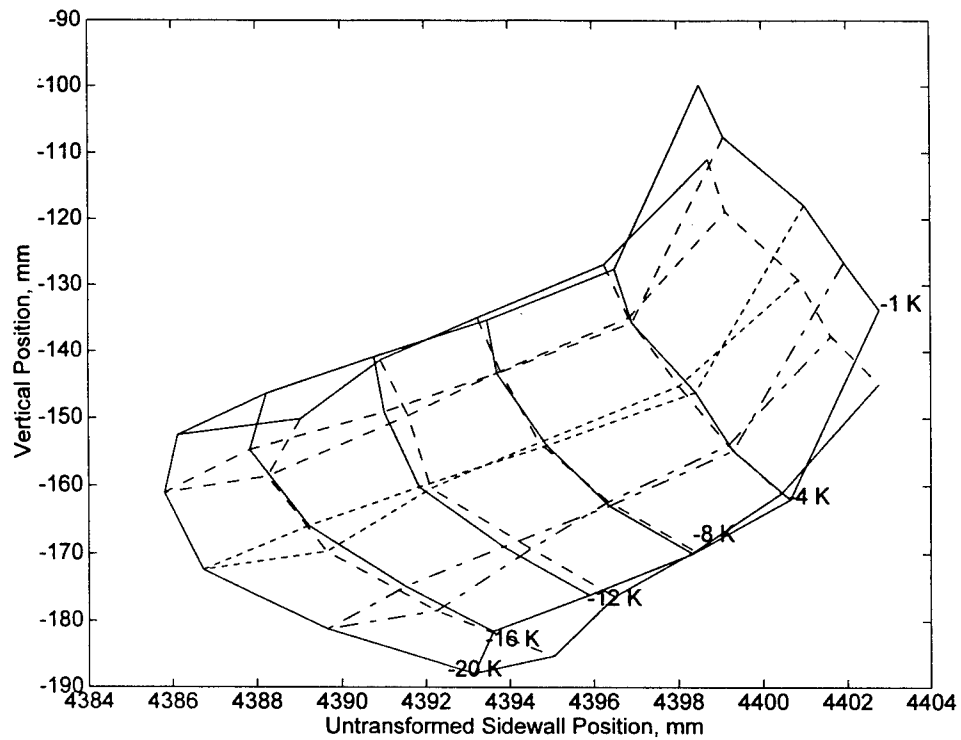


Figure 11. F-16 tire untransformed sidewall deformation under vertical test.

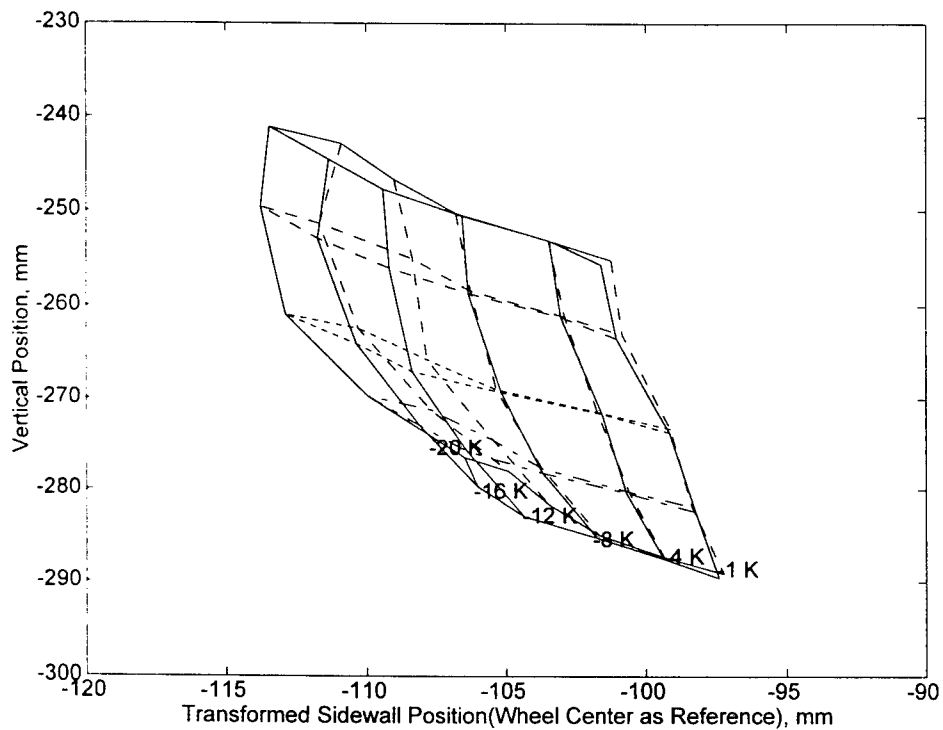


Figure 12. F-16 tire transformed sidewall deformation under vertical test.

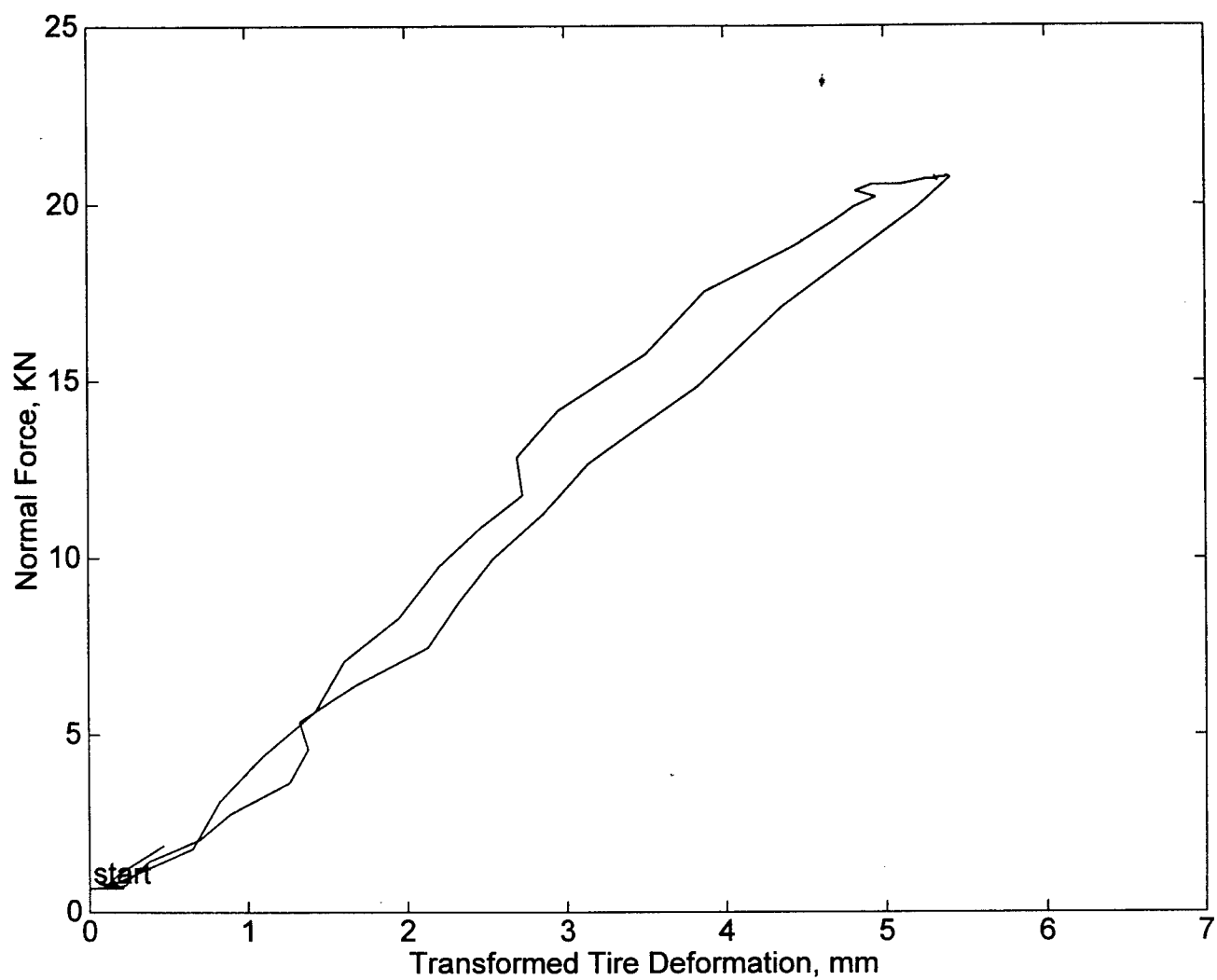


Figure 13. The relation between normal loading and vertical tire deflection for F-16 tire.

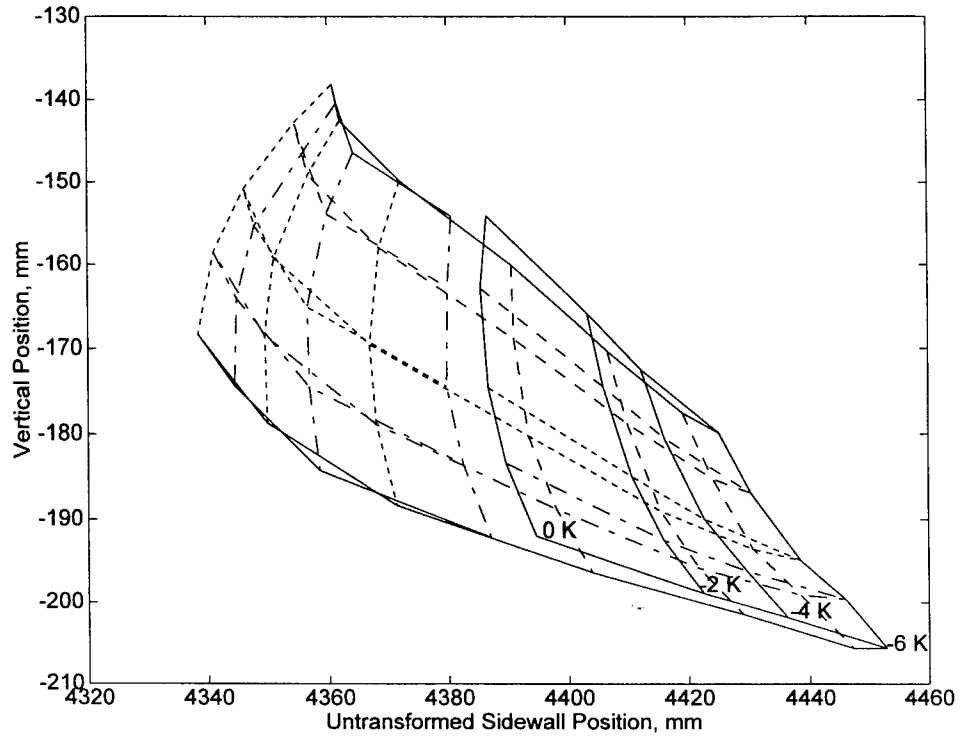


Figure 14. F-16 tire untransformed sidewall deformation under lateral-relaxation test.

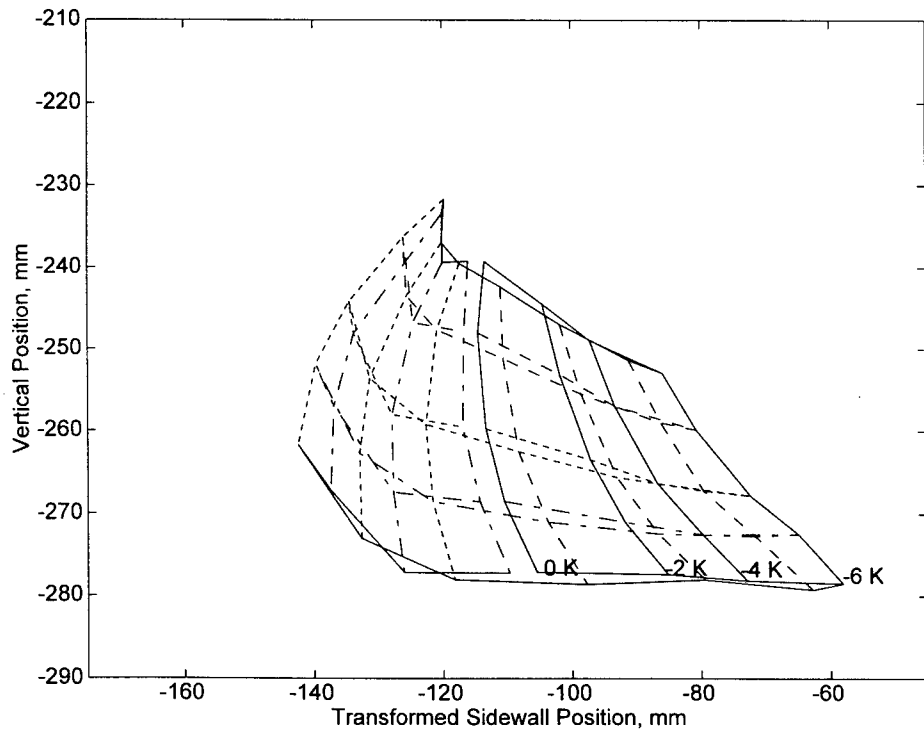


Figure 15. F-16 tire transformed sidewall deformation under lateral-relaxation test.

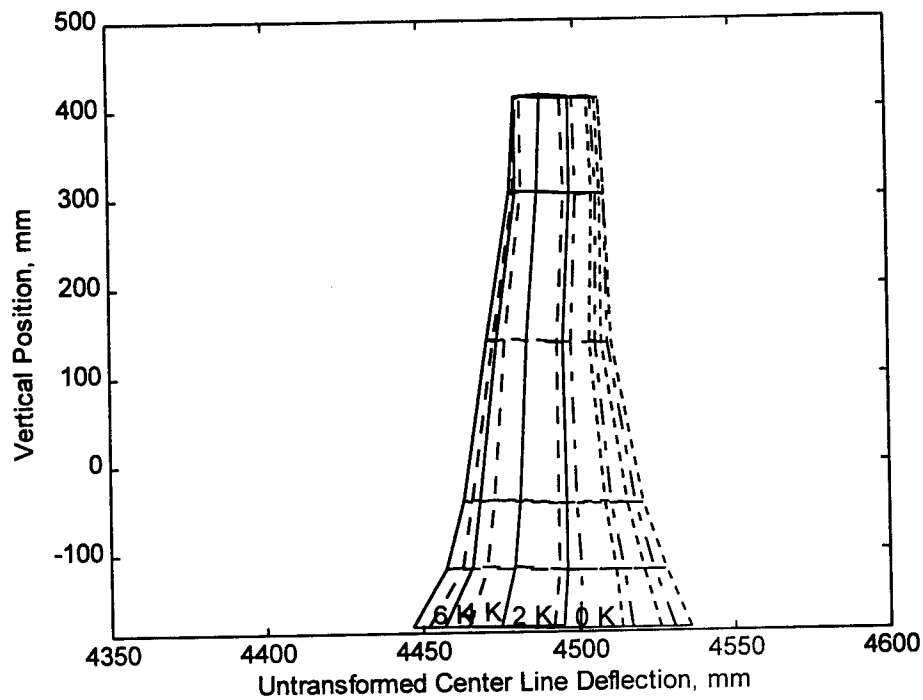


Figure 16. F-16 tire untransformed center line (periphery) deflection under lateral-relaxation test.

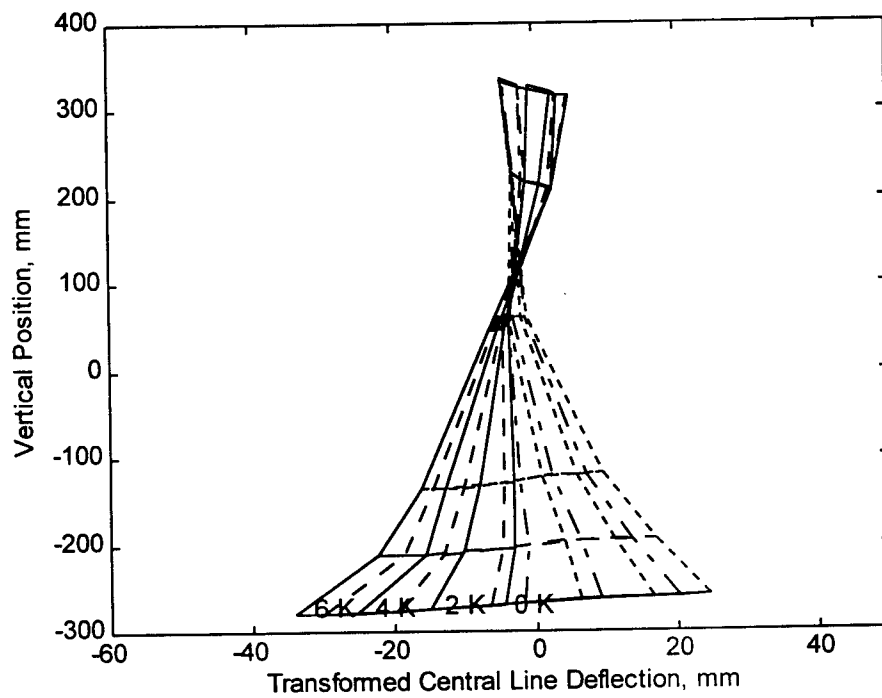


Figure 17. F-16 tire transformed center line (periphery) deflection under lateral-relaxation test.

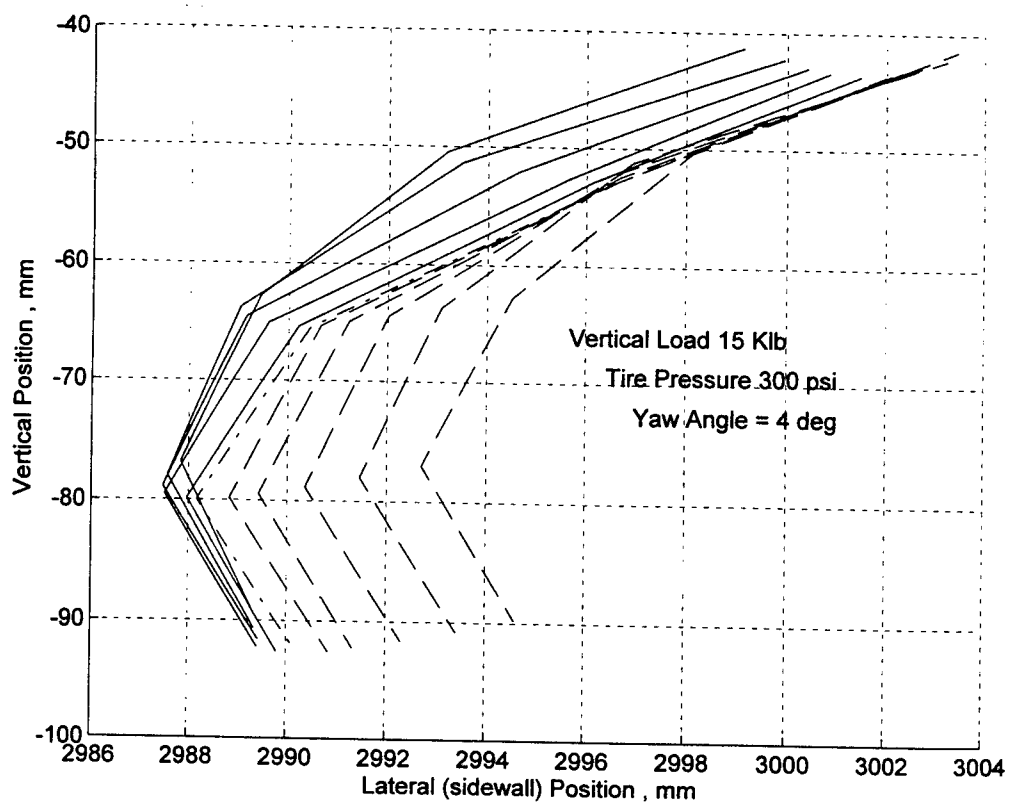


Figure 18. F-16 tire untransformed sidewall deformation under 4°-yaw free rolling test (normal load = 15kip).

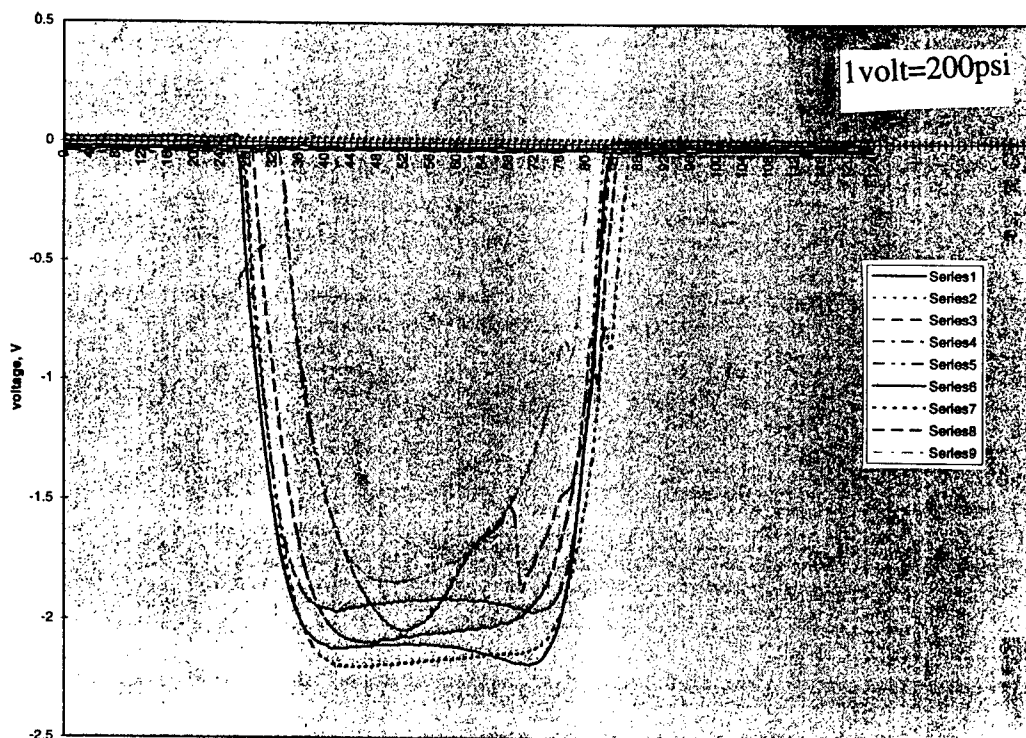


Figure 19. F-16 tire contact pressure under 4°-yaw free rolling test (normal load = 15kip).

Control of Linear Systems with Rate Limited Actuators and Its Applications to Flight Control Systems

Zongli Lin

Assistant Professor

Department of Applied Mathematics and Statistics

State University of New York at Stony Brook

Stony Brook, NY 11794-3600

Final Report for

AFOSR Summer Faculty Research Program

Wright Laboratory/Control and Dynamics Directorate

Wright-Patterson Air Force Base, OH 45433-7531

Sponsored by

Air Force Office of Scientific Research

Bolling Air Force Base, Washington, D.C.

and

Wright Laboratory

December 1997

Control of Linear Systems with Rate Limited Actuators and Its Applications to Flight Control Systems

Zongli Lin

Department of Applied Mathematics and Statistics
State University of New York at Stony Brook

Abstract

A design technique is proposed for obtaining feedback laws that cause the output of a linear system subject to actuator rate saturation to track a desired command signal. This design technique combines two design techniques recently developed for linear systems with position limited actuators, piecewise-linear LQ control ([20]) and low-and-high gain feedback ([18]), and hence takes advantage of both design techniques, while avoiding their disadvantages. In the case that only the output is available for feedback, the performance of the state feedback law is preserved by the use of a fast observer. The proposed design technique is also implemented in the Matlab environment. An open-loop exponentially unstable fighter aircraft is used to demonstrate the effectiveness of the proposed control design method.

Control of Linear Systems with Rate Limited Actuators and Its Applications to Flight Control Systems

Zongli Lin

1 Introduction

Every physical actuator is subject to constraints. These constraints include both position and rate saturation. In the past few years there has been much interest concerning the stabilization of linear systems with position saturating actuators, resulting in several promising design techniques. In this work, we will recourse to the low-and-high gain (LHG) design technique [18] and the piecewise linear LQ control (PLC) design technique [20]. Design techniques that are conceptually similar to the PLC design, the high performance bounded control (HPB), the scheduled low gain feedback and scheduled low-and-high gain feedback were recently developed in [15], [14] and [7], respectively. Additional related work might be found in the above mentioned papers and in [2].

While actuator position saturation has been addressed in the recent literature, few design techniques are currently available to deal with actuator rate saturation, which often presents a more serious challenge to control engineers, especially flight control engineers. It is known ([1]) that actuator rate saturation could induce a considerable phase-lag, which when associated with rate saturation has a destabilizing effect. For example, investigators have identified rate saturation as a contributing factor to the recent mishaps of the YF-22 [3] and the Gripen [5] prototypes and the first production Gripen [17]. For further discussion on the destabilizing effect of actuator rate saturation, see [1].

As a result, while visiting Wright Laboratory as a 1996 AFOSR summer faculty research associate, the PI and his lab. focus point at Wright Laboratory, Dr. Siva Banda have chosen the following topic to work on: control of linear systems with position and/or rate limited actuators. The PI's participation in the AFOSR Summer Faculty Research Program has been very fruitful. In what follows, we briefly summarize the results obtained during this Summer Faculty Research Program and the subsequent Summer Research Extension Program as follows.

⁰Wright Fellow, Chief of WL/FIGC (Flight Control Division, Flight Dynamics Directorate), Wright Laboratory, Wright-Patterson Air Force Base, Dayton, Ohio.

1. A composite nonlinear feedback design based on a nominal linear feedback is proposed for (exponentially unstable) second order linear systems with position limited actuators. The design yields nonlinear state feedback laws that both increase the speed of closed-loop system response to the command input and reduce the overshoot, while not imperiling the performance achieved by the nominal linear feedback controller in the face of actuator position saturation. As a result, arbitrarily fast step response is achieved with arbitrarily small overshoot. A flight control system is used to demonstrate the applicability and effectiveness of the design.

Further details on this result can be found in [8].

2. A robust stabilizing state feedback design technique for (exponentially unstable) linear systems with rate limited actuators is proposed. The design combines two design techniques recently developed for linear systems with position limited actuators, piecewise-linear LQ control [20] and low-and-high gain feedback [18]. An F-16 class open-loop unstable fighter aircraft model is used to demonstrate the applicability and effectiveness of the proposed design method. The combined design takes advantages of both design techniques, while avoiding their disadvantages. As a result, the resulting nonlinear feedback law achieves fast transient response and disturbance rejection in the face of actuator uncertainties. Moreover, the design also guarantees a fixed domain of attraction, which does not shrink as the values of the design parameters are increased to achieve higher degree of disturbance rejection.

Further details on this result can be found in [9].

3. It is shown that, if a linear system is asymptotically null controllable with bounded controls, then, when subject to both actuator position and rate saturation, it is semi-globally stabilizable by linear state feedback. If, in addition, the system is also detectable, then it is semi-global stabilizable via linear output feedback.

Further details on this result can be found in [6].

4. The design technique proposed in [9] (as described in Item 2) for stabilization of a linear system with rate-limited actuators is utilized to design feedback laws that cause the system output to track a desired command signal. The tracking control law is robust in the sense that it rejects the disturbance associated with the actuators in the face of actuator uncer-

tainties. An open-loop exponentially unstable fighter aircraft is used to demonstrate the effectiveness of the proposed control design method.

A paper that describes the state feedback result has been presented at 1997 AIAA GNC Conference [10]. A complete version of the paper that includes both state feedback and output feedback results has been written [11] and will be included in this final report.

5. Software in MATLAB is developed for the implementation and simulation of the results obtained.

Due to space limitation, in this final report, we only include detailed results that have not been published, i.e., [11]. More specifically, in [9] we proposed a method of designing stabilizing feedback control laws for linear systems taking into account the effect of actuator rate saturation ([9]). The proposed feedback design, referred to as the combined PLC/LHG state feedback design method, combines the two recently developed design techniques, the PLC and the LHG design techniques, both originally proposed for linear systems with actuator position saturation. It inherits the advantages of both design techniques, while avoiding their disadvantages. Thus, exact knowledge of the dynamics of the actuators will not be needed and the actuator disturbances can be rejected. In particular, in the LHG design [18], a low gain feedback law is first designed in such a way that the actuator does not saturate in position and the the closed-loop system remains linear. The gain is chosen low to enlarge the region in which the closed-loop system remains linear and hence to enlarge the domain of attraction of the closed-loop system. Then, utilizing an appropriate Lyapunov function for the closed-loop system under this low gain feedback control law, a linear high gain feedback control law is constructed and added to the low gain feedback control to form the final LHG feedback control law. Such a linear low-and-high gain feedback control law speeds up the transient response for the state in a certain subspace of the state space and is able to stabilize the system in the presence of input-additive plant uncertainties and to reject arbitrarily large bounded input-additive disturbances. The disadvantage of this control law is that the transient response for the state outside that subspace of the state space remains that of the low gain feedback, which is typically sluggish (due to low feedback gain for a large domain of attraction). On the other hand, the aim of the PLC [20] (HPB control [15]) or the scheduled low gain feedback design [14] schemes is to increase the feedback gain either piecewisely or continuously while adhering to the input bound as the trajectories converge toward the origin.

Such a design results in fast transient speed for all states. However, it lacks robustness to large uncertainties and the ability to reject disturbances.

Our objective is to utilize the combined PLC/LHG state feedback design technique developed in [9] for stabilization to design state feedback laws that achieve robust tracking performance. The signals to be tracked are modeled as the state of a reference system and the tracking control problem is formulated as an output regulation problem ([4], [13]). In the case that only the output is available for feedback, the performance of the state feedback laws is preserved by the use of a fast observer. We note here that the PLC design method as originally proposed in [20] is only for state feedback. Although the LHG output feedback design is given in [18], its design and analysis do not carry through in the combined PLC/LHG design, since the closed-loop system under the combined PLC/LHG state feedback law is discontinuous. Indeed, discontinuity in the PLC/LHG state feedback laws is the primary source of difficulty in its implementation with observer state.

The remainder of this report is organized as follows. In Section 2, we precisely formulate the two problems, the state feedback robust tracking problem and the output feedback robust tracking problem. Section 3 recalls from [9] the combined PLC/LHG state feedback design algorithm that was proposed for constructing stabilizing state feedback laws. This design algorithm is used in Section 4 to construct feedback laws that solve the state feedback robust tracking control problem as formulated in Section 2. Section 5 deals with the case that only output is available for feedback. In this case, a fast observer is used to implement the state feedback laws constructed in Section 4 and shown to preserve performance of the state feedback laws. In Section 6, a fighter aircraft model is used to demonstrate the effectiveness of the proposed design algorithm. Concluding remarks are made in Section 7.

2 Problem Formulation

Consider the linear dynamical system

$$\begin{cases} \dot{x} = Ax + Bv, & x(0) \in \mathcal{X} \subset \mathbb{R}^n \\ \dot{v} = \text{sat}_\Delta(-Tv + Tu + d), & v(0) \in \mathcal{V} \subset \mathbb{R}^m \\ \dot{w} = Sw, & w(0) \in \mathcal{W} \subset \mathbb{R}^s \\ e = Cx + Qw \\ y = Ex \end{cases} \quad (1)$$

where the first equation describes the plant, with state $x \in \mathbb{R}^n$, input $u \in \mathbb{R}^m$ and output $y \in \mathbb{R}^q$, the second equation represents actuator dynamics with state $v \in \mathbb{R}^m$ and is subject to disturbance $d \in \mathbb{R}^m$ and rate saturation $\text{sat}_\Delta(\cdot)$, the third equation defines the reference signal $-Qw \in \mathbb{R}^p$ to be tracked by the plant output $Cx \in \mathbb{R}^p$. For $\Delta = (\Delta_1, \Delta_2, \dots, \Delta_m)$, $\Delta_i > 0$, the actuator rate saturation function $\text{sat}_\Delta : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is the standard saturation function, i.e., $\text{sat}_\Delta(v) = [\text{sat}_{\Delta_1}(v_1), \text{sat}_{\Delta_2}(v_2), \dots, \text{sat}_{\Delta_m}(v_m)]$, $\text{sat}_{\Delta_i}(v_i) = \text{sign}(v_i) \min\{\Delta_i, |v_i|\}$, the positive definite diagonal matrices $T = \text{diag}(\tau_1, \tau_2, \dots, \tau_m)$ represents the “time constants” of the actuators. Finally, all three sets \mathcal{X} , \mathcal{V} and \mathcal{W} are bounded.

We also make the following assumptions on the system.

Assumption 2.1 *The pair (A, B) is stabilizable;*

Assumption 2.2 *The eigenvalues of S have nonnegative real parts.*

Assumption 2.3 *The disturbance is uniformly bounded by a known (arbitrarily large) constant D , i.e., $\|d(t)\| \leq D, \forall t \geq 0$.*

Remark 2.4 *We note that Assumption 2.1 is obviously necessary, Assumption 2.2 does not involve any loss of generality since asymptotically stable modes of the reference model do not affect the tracking performance of the plant, and Assumption 2.3 is satisfied by essentially all practical disturbances.*

Before stating the problem, we make the following preliminary definition.

Definition 2.5 *The data $(D, \mathcal{Z}_\infty, \mathcal{E}_\infty)$ is said to be admissible for state feedback [output feedback] if D is a nonnegative real number, and \mathcal{Z}_∞ and \mathcal{E}_∞ are respectively a subset of \mathbb{R}^{n+m} [\mathbb{R}^{2n+m}] and \mathbb{R}^p both containing the origin as an interior point.*

The objective is to provide a design algorithm for solving the following problems.

Problem 1 [State feedback robust tracking problem.] *Given the data $(D, \mathcal{Z}_\infty, \mathcal{E}_\infty)$, admissible for state feedback, find, if possible, a state feedback law $u = F(x, v, w)$, such that the closed-loop system satisfies,*

1. *Every trajectory of the system*

$$\begin{cases} \dot{x} = Ax + Bv \\ \dot{v} = \text{sat}_{\Delta}(-Tv + TF(x, v, 0) + d) \end{cases} \quad (2)$$

starting from $\mathcal{X} \times \mathcal{V}$ enters \mathcal{Z}_{∞} in a finite time and remains in it thereafter.

2. *For all $x(0) \in \mathcal{X}$, $v(0) \in \mathcal{V}$ and $w(0) \in \mathcal{W}$, the solution of the closed-loop system is such that $e(t)$ enters the set \mathcal{E}_{∞} in a finite time and remains in it thereafter.*

Problem 2 [Output feedback robust tracking problem.] *Given the data $(D, \mathcal{Z}_{\infty}, \mathcal{E}_{\infty})$, admissible for output feedback, find, if possible, an output feedback law*

$$\begin{cases} \dot{\psi} = \alpha(\psi, y), \quad \psi(0) \in \mathcal{X} \subset \mathbb{R}^n \\ u = F(\psi, v, w) \end{cases}$$

such that the closed-loop system satisfies

1. *Every trajectory of the system*

$$\begin{cases} \dot{x} = Ax + Bv \\ \dot{v} = \text{sat}_{\Delta}(-Tv + TF(\psi, v, 0) + d) \\ \dot{\psi} = \alpha(\psi, Ex) \end{cases}$$

starting from $\mathcal{X} \times \mathcal{V} \times \mathcal{X}$ enters \mathcal{Z}_{∞} in a finite time and remains in it thereafter.

2. *For all $x(0), \psi(0) \in \mathcal{X}$, $v(0) \in \mathcal{V}$ and $w(0) \in \mathcal{W}$, the solution of the closed-loop system is such that $e(t)$ enters the set \mathcal{E}_{∞} in a finite time and remains in it thereafter.*

3 A Combined PLC/LHG Design Algorithm

In this section, we recall from [9] a design algorithm that is a combination of the piecewise linear LQ control [20] and the low-and-high gain feedback [18]. For completeness, we organize this section as follows. Subsections 3.1 and 3.2 respectively recapitulate the PLC and the LHG design techniques. Subsection 3.3 presents the combined PLC/LHG design algorithm.

3.1 Piecewise Linear LQ Control Design

Consider the linear dynamical system subject to actuator position saturation,

$$\dot{x} = Ax + B\text{sat}_{\Delta}(u), \quad x(0) \in \mathcal{X} \subset \mathbb{R}^n, \quad u \in \mathbb{R}^m \quad (3)$$

where the saturation function $\text{sat}_\Delta : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is as defined in Section 2, the pair (A, B) is assumed to be stabilizable, and \mathcal{X} is bounded.

The PLC design is based on the following LQ algebraic Riccati equation,

$$A'P + PA - PBR^{-1}B'P + I = 0 \quad (4)$$

where $R = \text{diag}(\epsilon) = \text{diag}(\epsilon_1, \epsilon_2, \dots, \epsilon_m)$, $\epsilon_i > 0$, are the design parameters to be chosen later.

Key to the PLC scheme is the notion of invariant sets. A nonempty subset of ϵ in \mathbb{R}^n is positively invariant if for a dynamical system and for any initial condition $x(0) \in \epsilon$, $x(t) \in \epsilon$ for all $t \geq 0$. For the closed-loop system comprising system (3) and the LQ control $u = -R^{-1}B'Px$, simple Lyapunov analysis shows that the Lyapunov level set

$$\epsilon(P, \rho) = \{x : x'Px \leq \rho\}, \quad \forall \rho > 0$$

is an invariant set, provided that saturation does not occur for all $x \in \epsilon(P, \rho)$. To avoid the saturation from occurring, while fully utilizing the available control capacity, for a given ρ , $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_m)$ will be chosen to be the smallest such that

$$|u_i| = \left| \frac{1}{\epsilon_i} B_i' Px \right| \leq \Delta_i, \quad \forall x \in \epsilon(P, \rho)$$

where B_i is the i th column of matrix B and u_i is the i th element of u . The existence and uniqueness of such an ϵ are established, and an algorithm for computing such an ϵ is also given, in [20]. More specifically, it is shown through the existence of a unique fixed point that the following iteration converges from any initial value to the desired value of ϵ ,

$$\epsilon_{n+1} = \sqrt{\rho} \Phi(\epsilon_n) \quad (5)$$

where

$$\Phi(\epsilon) = [\phi_1(\epsilon), \phi_2(\epsilon), \dots, \phi_m(\epsilon)]'$$

and for each $i = 1$ to m ,

$$\phi_i(\epsilon) = \frac{1}{\Delta_i} \sqrt{B_i' P(\epsilon) B_i}$$

The aim of the PLC scheme is to increase the state feedback gain piecewisely while adhering to actuator bounds as the trajectories converge towards the origin. This is achieved by constructing nested level sets, $\epsilon_0, \epsilon_1, \dots, \epsilon_N$, in such a way that the trajectories traverse successively the surface of each ϵ_i and the control law is switched to higher and higher gains as each surface is crossed.

The procedure for designing a PLC law is as follows. Given the set of initial conditions $\mathcal{X} \subset \mathbb{R}^n$, choose an initial level set ε_0 as,

$$\varepsilon_0 = \inf_{\rho} \{\varepsilon(P, \rho) : \mathcal{X} \subset \varepsilon(P, \rho)\} \quad (6)$$

We denote the value of ρ associated with ε_0 as ρ_0 , and the corresponding values of ϵ , R and P as ϵ_0 , R_0 and P_0 respectively. A simple approach to determining ε_0 and ρ_0 can also be found in [20]. More specifically, it is shown that the size of ε_0 grows monotonically as the parameter ρ grows. Hence, ε_0 and ρ_0 can be determined by a simple iteration procedure. Here we would like to note that, as explained in [20], increasing ρ indefinitely for exponentially unstable A will not result in an ε_0 that grows without bound.

To determine the inner level sets ε_i 's, choose successively smaller ρ_i where $\rho_{i+1} < \rho_i$ for each $i = 1, 2, \dots, N$. A simple choice of such ρ_i 's is the geometric sequence of the form

$$\rho_i = \rho_0(\Delta\rho)^i, \quad i = 0, 1, 2, \dots, N$$

where the ' ρ -reduction' factor $\Delta\rho \in (0, 1)$. (Consequently, the values of ϵ , R and P associated with each of these ρ_i 's are denoted as ϵ_i , R_i and P_i respectively.) For a discussion on the choice of N and $\Delta\rho$, see [20].

As shown in [20], a critical property of such a sequence of level sets ε_i is that they are strictly nested in the sense that ε_{i+1} is strictly inside ε_i for each $i = 0$ to $N - 1$. Here and throughout this report, by saying set S_2 is strictly inside set S_1 we mean that $S_2 \subset S_1$ and their boundaries do not overlap.

3.2 Low-and-High Gain Feedback Design

Consider the linear dynamical system subject to actuator position saturation, input additive disturbances and uncertainties,

$$\dot{x} = Ax + B\text{sat}_{\Delta}(u + f(x) + d), \quad x(0) \in \mathcal{X} \subset \mathbb{R}^n, u \in \mathbb{R}^m \quad (7)$$

where the saturation function $\text{sat}_{\Delta} : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is as defined in Section 2, the locally Lipschitz function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ represents the input additive plant uncertainties and d the input-additive disturbance. The LHG feedback design for this system is given as follows. First, the level set

ε_0 is determined as in the PLC design. Correspondingly, a state feedback law with possibly low feedback gain is determined as,

$$u_L = -R_0^{-1}B'P_0x$$

A high gain state feedback is then constructed as,

$$u_H = -kR_0^{-1}B'P_0x, \quad k \geq 0$$

The final low-and-high gain state feedback is then given by a simple addition of the low and high gain feedbacks u_L and u_H , viz.,

$$u = -(1+k)R_0^{-1}B'P_0x, \quad k \geq 0$$

Here the design parameter k is referred to as the high gain parameter. As demonstrated in [18] the freedom in choosing the value of this high gain parameter can be utilized to achieve robust stabilization in the presence of input additive plant uncertainties $f(x)$ and input-additive disturbance rejection. Moreover, the transient speed for the states not in the range space of $B'P_0$ will increase as the value of k increases. To see this, let us consider the following Lyapunov function,

$$V_0(x) = x'P_0x$$

The evaluation of \dot{V} along the trajectories of the closed-loop system in the absence of uncertainties and disturbances gives,

$$\begin{aligned} \dot{V} &= -x'x - x'P_0BR_0^{-1}B'P_0x + 2x'P_0B[\text{sat}_\Delta(-(k+1)R_0^{-1}B'P_0x) + R_0^{-1}B'P_0x] \\ &= -x'x - x'P_0BR_0^{-1}B'P_0x - 2 \sum_{i=1}^m v_i[\text{sat}_{\Delta_i}((k+1)v_i) - v_i] \end{aligned}$$

where we have denoted the i th element of $v = -R_0^{-1}B'P_0x$ as v_i . By the choice of P_0 , it is clear that $|v_i| \leq \Delta_i$ and hence $-v_i[\text{sat}_{\Delta_i}((k+1)v_i) - v_i] \leq 0$, for each $i = 1$ to m . For any x not in the range space of $B'P_0$, that is $B'P_0x \neq 0$, then, for any i such that $v_i \neq 0$, $-v_i[\text{sat}_{\Delta_i}((k_2+1)v_i) - v_i] < -v_i[\text{sat}_{\Delta_i}((k_1+1)v_i) - v_i]$ if $k_2 > k_1$. However, for any x in the null space of $B'P_0$, $-v_i[\text{sat}_{\Delta_i}((k+1)v_i) - v_i] = 0$ for any value of k .

3.3 Combined PLC/LHG Feedback Design

In this subsection, we present the proposed combined PLC/LHG state feedback design for linear systems subject to actuator rate saturation (1).

The feedback control law design is carried out in the following three steps.

Step 1. Choose a pre-feedback

$$u = v + \bar{u} \quad (8)$$

Let $\tilde{x} = [x', v']'$. Then the system (1) under the above pre-feedback is given by,

$$\dot{\tilde{x}} = \tilde{A}\tilde{x} + \tilde{B}\text{sat}_\Delta(T\bar{u} + d), \quad \tilde{x}(0) \in \mathcal{X} \times \mathcal{V} \subset \mathbb{R}^{n+m} \quad (9)$$

where

$$\tilde{A} = \begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} 0 \\ I \end{bmatrix} \quad (10)$$

Assumption 2.1, i.e., the pair (A, B) is stabilizable, implies that (\tilde{A}, \tilde{B}) is stabilizable.

Step 2. Apply the PLC state feedback design algorithm to system (9), and obtain a sequence of nested level sets $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_N$ (and correspondingly, the parameters $\epsilon_0, \epsilon_1, \dots, \epsilon_N$) and a piecewise linear state feedback law,

$$\bar{u} = \begin{cases} \bar{u}_i = -T^{-1}\tilde{R}_i^{-1}\tilde{B}'\tilde{P}_i\tilde{x} & \text{for } \tilde{x} \in \varepsilon_i \setminus \varepsilon_{i+1}, i = 0, 1, \dots, N-1 \\ \bar{u}_N = -T^{-1}\tilde{R}_N^{-1}\tilde{B}'\tilde{P}_N\tilde{x} & \text{for } \tilde{x} \in \varepsilon_N \end{cases} \quad (11)$$

Step 3. Design the LHG state feedback based on the PLC feedback law (11) and obtain the following combined final PLC/LHG feedback law,

$$u = \begin{cases} u_i = v - (k+1)T^{-1}\tilde{R}_i^{-1}\tilde{B}'\tilde{P}_i\tilde{x} & \text{for } \tilde{x} \in \varepsilon_i \setminus \varepsilon_{i+1}, i = 0, 1, \dots, N-1 \\ u_N = v - (k+1)T^{-1}\tilde{R}_N^{-1}\tilde{B}'\tilde{P}_N\tilde{x} & \text{for } \tilde{x} \in \varepsilon_N \end{cases} \quad (12)$$

where $k \geq 0$ is a design parameter to be specified later.

4 Robust Tracking via State Feedback

In this section, we will show that the combined PLC/LHG design technique as recalled in the previous section can be used to construct feedback laws that solve the state feedback robust tracking problem (i.e., Problem 1) as formulated in Section 2. The state feedback results are presented in a theorem as follows.

Theorem 4.1 *Let Assumptions 2.1-2.3 hold, and given the data $(D, \mathcal{W}_\infty, \mathcal{E}_\infty)$ admissible for state feedback, then Problem 1 is solvable if there exist matrices Π and Γ such that*

1. *they satisfy the linear matrix equations,*

$$\begin{cases} \Pi S = \tilde{A}\Pi + \tilde{B}\Gamma \\ \tilde{C}\Pi + Q = 0 \end{cases} \quad (13)$$

where \tilde{A} and \tilde{B} are as given by (10), and $\tilde{C} = [C \ 0_{p \times m}]$.

2. *there exists a $\delta = [\delta_1, \delta_2, \dots, \delta_m]$, $\delta_i > 0$, such that $|\Gamma_i w(t)| \leq \Delta_i - \delta_i$ for all $w(0) \in \mathcal{W}$ and all $t \geq 0$, where Γ_i is the i th row of Γ .*

Moreover, the feedback laws that solve Problem 1 can be explicitly constructed using the combined PLC/LHG design technique as recalled in the previous section.

Remark 4.2 *It follows from [4] that Condition 1 is necessary for solving Problem 1, while, similar arguments as in [13] show that, under some extra mild conditions on the plant and the reference model, Condition 2 is also necessary.*

Proof. We will prove this theorem by explicitly constructing a family of combined PLC/LHG state feedback laws, parameterized in a parameter k and showing that there exists a $k^* > 0$ such that for any $k \geq k^*$, the feedback law solves Problem 1.

Construction of Parameterized State Feedback Laws

Step 1. Carry out Step 1 of the combined PLC/LHG feedback design of Section 3.3 as follows. Choose a pre-feedback

$$u = v + \bar{u} \quad (14)$$

Let $\tilde{x} = [x', v']'$, Then system (1) under the above pre-feedback becomes,

$$\dot{\tilde{x}} = \tilde{A}\tilde{x} + \tilde{B}\text{sat}_\Delta(T\bar{u} + d) \quad (15)$$

where \tilde{A} and \tilde{B} are given by (10). Also let $\tilde{x}(0) \in \Xi = (\mathcal{X} \times \mathcal{V}) \cup \{\tilde{x} - \Pi w : \tilde{x} \in \mathcal{X} \times \mathcal{V}, w \in \mathcal{W}\}$.

Step 2. Carry out Step 2 of the combined PLC/LHG feedback design of Section 3.3 using δ_i instead of Δ_i and Ξ instead of $\mathcal{X} \times \mathcal{V}$, and obtain a sequence of nested level sets $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_N$ (and correspondingly, the parameters $\epsilon_0, \epsilon_1, \dots, \epsilon_N$) and a piecewise linear feedback law,

$$\bar{u} = \begin{cases} \bar{u}_i = -T^{-1}\tilde{R}_i^{-1}\tilde{B}'\tilde{P}_i\tilde{x} & \text{for } \tilde{x} \in \varepsilon_i \setminus \varepsilon_{i+1}, i = 0, 1, \dots, N-1 \\ \bar{u}_N = -T^{-1}\tilde{R}_N^{-1}\tilde{B}'\tilde{P}_N\tilde{x} & \text{for } \tilde{x} \in \varepsilon_N \end{cases} \quad (16)$$

Step 3. Design the final combined PLC/LHG tracking control laws as,

$$u = \begin{cases} u_i = v - (k+1)T^{-1}\tilde{R}_i^{-1}\tilde{B}'\tilde{P}_i\tilde{x} + T^{-1}[(k+1)\tilde{B}'\tilde{R}_i^{-1}\tilde{P}_i\Pi + \Gamma]w, & \text{for } \tilde{x} \in \varepsilon_i \setminus \varepsilon_{i+1}, i = 0, 1, \dots, N-1 \\ u_N = v - (k+1)T^{-1}\tilde{R}_N^{-1}\tilde{B}'\tilde{P}_N\tilde{x} + T^{-1}[(k+1)\tilde{B}'\tilde{R}_N^{-1}\tilde{P}_N\Pi + \Gamma]w & \text{for } \tilde{x} \in \varepsilon_N \end{cases} \quad (17)$$

where $k \geq 0$ is a design parameter to be specified later.

We now proceed to show that,

Point 1. there exists a $k_1^* > 0$ such that, for all $k \geq k_1^*$, Item 1 of Problem 1 holds;

Point 2. there exists a $k_2^* > 0$ such that, for all $k \geq k_2^*$, Item 2 of Problem 1 holds,

from which the result of Theorem 4.1 follows with $k^* = \max\{k_1^*, k_2^*\}$.

The existence of such a k_1^* follows readily from [9]. It remains to show the existence of $k_2^* > 0$.

To do so, let us introduce an invertible, triangular coordinate change $\xi = \tilde{x} - \Pi w$. Using Condition 1 of the theorem, we have,

$$\dot{\xi} = \tilde{A}\xi + \tilde{B}[\text{sat}_{\Delta}(-Tv + Tu + d) - \Gamma w] \quad (18)$$

where u is given by (17). Let $\Xi_{\infty} \subset \mathbb{R}^{n+m}$ be such that $\xi \in \Xi_{\infty}$ implies that $C\xi \in \mathcal{E}_{\infty}$. We next show that there exists a $k_2^* > 0$ such that for any $k \geq k_2^*$ the solution of (18) starting from any $\xi(0) \in \Xi$ enters the set Ξ_{∞} in a finite time and remains in it thereafter. This can be done in two steps. In the first step, we show that, for each $i = 0$ to $N-1$, there exists a $k_{2i}^* > 0$, such that for all $k \geq k_{2i}^*$, in the presence of any d satisfying Assumption 2.3, all trajectories starting from $\varepsilon_i \setminus \varepsilon_{i+1}$ will remain in ε_i and enter into the inner level set ε_{i+1} in a finite time. This in turn implies that, for any $k \geq \max\{k_{20}^*, k_{21}^*, \dots, k_{2N-1}^*\}$, all the trajectories of the closed-loop system starting from $\tilde{x} \in \varepsilon_0$ will enter the inner-most level set ε_N in a finite time. The second step of the proof is to show that there exists a $k_{2N}^* > 0$ such that, for all $k \geq k_{2N}^*$, all the trajectories of the closed-loop system starting from ε_N will remain in it and enter and remain in the set Ξ_{∞} in a finite time. Once these two steps are completed, the proof of Point 2 is then completed by taking $k_2^*(D, \Xi_{\infty}) = \max\{k_{20}^*, k_{21}^*, \dots, k_{2N}^*\}$.

We start by considering the closed-loop system (18) for $\xi \in \varepsilon_i \setminus \varepsilon_{i+1}$, $i = 0$ to N ,

$$\begin{aligned} \dot{\xi} &= \tilde{A}\xi + \tilde{B}[\text{sat}(-(k+1)\tilde{R}_i^{-1}\tilde{B}'\tilde{P}_i\xi + \Gamma w + d) - \Gamma w] \\ &= (\tilde{A} - \tilde{B}\tilde{R}_i^{-1}\tilde{B}'\tilde{P}_i)\xi + \tilde{B}[\text{sat}_{\Delta}(-(k+1)\tilde{R}_i^{-1}\tilde{B}'\tilde{P}_i\xi + \Gamma w + d) + \tilde{R}_i^{-1}\tilde{B}'\tilde{P}_i\xi - \Gamma w] \end{aligned} \quad (19)$$

where $\varepsilon_{N+1} = \emptyset$.

Pick the Lyapunov function,

$$V_i = \xi' \tilde{P}_i \xi \quad (20)$$

The evaluation of \dot{V}_i along the trajectories of the closed-loop system (19) gives,

$$\begin{aligned} \dot{V}_i &= -\xi' \xi - \xi' \tilde{P}_i \tilde{B} \tilde{R}_i^{-1} \tilde{B}' \tilde{P}_i \xi + 2\xi' \tilde{P}_i \tilde{B} [\text{sat}_{\Delta}(-(k+1)\tilde{R}_i^{-1} \tilde{B}' \tilde{P}_i \xi + \Gamma w + d) + \tilde{R}_i^{-1} \tilde{B}' \tilde{P}_i \xi - \Gamma w] \\ &\leq -\xi' \xi - 2 \sum_{i=1}^m v_i [\text{sat}_{\Delta_i}((k+1)v_i + \theta_i + d_i) - v_i - \theta_i] \end{aligned}$$

where we have denoted the i th elements of $v = -\tilde{R}_i^{-1} \tilde{B}' \tilde{P}_i \xi$, Γw and d respectively as v_i , θ_i and d_i .

By the construction of ε_i , it is clear that $|v_i + \theta_i| \leq \Delta_i$ for all $\xi \in \varepsilon_i$. Hence we have,

$$|kv_i| \geq |d_i| \implies v_i [\text{sat}_{\Delta_i}((k+1)v_i + \theta_i + d_i) - v_i - \theta_i] \geq v_i [\text{sat}_{\Delta_i}(v_i + \theta_i) - v_i - \theta_i] = 0$$

and,

$$|kv_i| < |d_i| \implies v_i [\text{sat}_{\Delta_i}((k+1)v_i + \theta_i + d_i) - v_i - \theta_i] \leq \frac{2|d_i|^2}{k}$$

where we have used the fact that,

$$|\text{sat}_{\Delta_i}(v_1) - \text{sat}_{\Delta_i}(v_2)| \leq |v_1 - v_2|, \quad \forall v_1, v_2 \in \mathbb{R}$$

Hence, we can conclude that, for all $\xi \in \varepsilon_i \setminus \varepsilon_{i+1}$, $i = 0$ to $N-1$,

$$\dot{V}_i \leq -\xi' \xi + \frac{4\|d\|^2}{k} \leq -\xi' \xi + \frac{4D^2}{k} \quad (21)$$

To complete the first step of the proof, for each $i = 0$ to $N-1$, we let

$$k_i^*(D) = \frac{5\lambda_{\max}(\tilde{P}_{i+1})D^2}{\rho_{i+1}}$$

We then have that, for all $k \geq k_i^*$,

$$\dot{V}_i \leq -\frac{1}{\lambda_{\max}(\tilde{P}_{i+1})} \left(V_{i+1} - \frac{4\lambda_{\max}(\tilde{P}_{i+1})D^2}{k} \right) \quad (22)$$

and hence,

$$\dot{V}_i < 0, \quad \forall \tilde{x} \in \varepsilon_i \setminus \varepsilon_{i+1} \quad (23)$$

which, together with the fact that ε_{i+1} is strictly inside ε_i , shows that all the trajectories of the closed-loop system starting from $\varepsilon_i \setminus \varepsilon_{i+1}$ will remain in ε_i and enter the level set ε_{i+1} in a finite time.

For the second step of the proof, let $\rho_{N+1} \in (0, \rho_N]$ be such that $\varepsilon(\tilde{P}_N, \rho_{N+1}) \subset \mathcal{E}_\infty$. The existence of such a ρ_{N+1} is due to the fact that Ξ_∞ contains the origin of the state space as an interior point. Choose $k_{2N}^*(D, \Xi_\infty)$ as follows,

$$k_N^*(D, \Xi_\infty) = \frac{5\lambda_{\max}(\tilde{P}_N)D^2}{\rho_{N+1}}$$

We will also have that, for all $k \geq k_{2N}^*$,

$$\dot{V}_N \leq -\frac{1}{\lambda_{\max}(\tilde{P}_N)} \left(V_N - \frac{4\lambda_{\max}(\tilde{P}_N)D^2}{k} \right) \quad (24)$$

and hence,

$$\dot{V}_N < 0, \quad \forall \xi \in \varepsilon_N \setminus \varepsilon(\tilde{P}_N, \rho_{N+1}) \quad (25)$$

which shows that all trajectories of the closed-loop system starting from ε_N will remain in ε_N and enter the set $\varepsilon(\tilde{P}_N, \rho_{N+1}) \subset \Xi_\infty$ in a finite time and remain in it thereafter.

So far, we have shown that any trajectory of the closed-loop system (19) starting from the set Ξ will enter the set Ξ_∞ in a finite time and remain in it thereafter. Noting Condition 1 of the theorem implies that $e = C\xi$, and recalling that $\xi \in \Xi_\infty$ implies that $e \in \mathcal{E}_\infty$, which concludes the proof of Theorem 4.1. \square

5 Robust Tracking via Output Feedback

In this section, we consider the case that only the output is available for feedback, i.e., the output feedback robust tracking problem. We will use a fast observer to preserve the performance of the combined PLC/LHG state feedback laws proposed in the previous section. In order to build the fast observer, we make the following assumption,

Assumption 5.1 *The pair (A, E) is observable.*

Our output feedback result is given as follows.

Theorem 5.2 *Let Assumptions 2.1-2.3 and 5.1 hold, and given the data $(D, \mathcal{W}_\infty, \mathcal{E}_\infty)$ admissible for output feedback, then Problem 2 is solvable if there exist matrices Π and Γ such that*

1. they satisfy the linear matrix equations,

$$\begin{cases} \Pi S = \tilde{A}\Pi + \tilde{B}\Gamma \\ \tilde{C}\Pi + Q = 0 \end{cases} \quad (26)$$

where \tilde{A} and \tilde{B} are as given by (10), and $\tilde{C} = [C \ 0_{p \times m}]$.

2. there exists a $\delta = [\delta_1, \delta_2, \dots, \delta_m]$, $\delta_i > 0$, such that $|\Gamma_i w(t)| \leq \Delta_i - \delta_i$ for all $w(0) \in \mathcal{W}$ and all $t \geq 0$, where Γ_i is the i th row of Γ .

Moreover, the output feedback laws that solve Problem 2 can be explicitly constructed by implementing the state feedback law of the previous section with a fast observer.

Proof: We will prove this theorem by explicitly constructing a family of fast observer based combined PLC/LHG output feedback laws, parameterized in two parameters k and ℓ and showing that, there exists a $k^* > 0$, and for each $k \geq k^*$, there exists an $\ell^*(k) > 0$ such that, for any $\ell \geq \ell^*(k)$, $k \geq k^*$, the output feedback law solves Problem 2.

Construction of Parameterized Output Feedback Laws

Step 1: Fast Observer Design. The family of fast observers, parameterized in ℓ is given by,

$$\dot{\hat{x}} = A\hat{x} + L(\ell)(y - C\hat{x}) + Bv, \quad \hat{x}(0) \in \mathcal{X} \quad (27)$$

where $L(\ell)$, $\ell > 0$ is chosen such that

$$\lambda(A + L(\ell)C) = \ell\Lambda \quad (28)$$

where Λ is any set of n complex numbers, with negative real parts and close under complex conjugation.

Step 2: State Feedback Design. Carry out the state feedback design as in the proof of Theorem 4. In the design, instead of letting $\Xi = (\mathcal{X} \times \mathcal{V}) \cup \{\tilde{x} - \Pi w : \tilde{x} \in \mathcal{X} \times \mathcal{V}, w \in \mathcal{W}\}$, let Ξ be any bounded set such that $(\mathcal{X} \times \mathcal{V}) \cup \{\tilde{x} - \Pi w : \tilde{x} \in \mathcal{X} \times \mathcal{V}, w \in \mathcal{W}\}$ is strictly inside it. Let the resulting state feedback laws be given by,

$$u = \begin{cases} u_i = v - (k+1)T^{-1}\tilde{R}_i^{-1}\tilde{B}'\tilde{P}_i\tilde{x} + T^{-1}[(k+1)\tilde{B}'\tilde{R}_i^{-1}\tilde{P}_i\Pi + \Gamma]w, \\ \quad \text{for } \tilde{x} \in \varepsilon_i \setminus \varepsilon_{i+1}, i = 0, 1, \dots, N-1 \\ u_N = v - (k+1)T^{-1}\tilde{R}_N^{-1}\tilde{B}'\tilde{P}_N\tilde{x} + T^{-1}[(k+1)\tilde{B}'\tilde{R}_N^{-1}\tilde{P}_N\Pi + \Gamma]w \quad \text{for } \tilde{x} \in \varepsilon_N \end{cases} \quad (29)$$

where

$$\varepsilon_i = \{\tilde{x} \in \mathbb{R}^{n+m} : \tilde{x}' \tilde{P}_i \tilde{x} \leq \rho_i\}, i = 0, 1, \dots, N$$

is the sequence of strictly nested level sets and $k \geq 0$ is a design parameter to be specified later.

Step 3: Output Feedback Laws. The final output feedback laws are obtained by implementing the state feedback law as obtained in Step 1 with the fast observer as obtained in Step 2. The resulting output feedback law is given by,

$$u = \begin{cases} u_i = v - (k+1)T^{-1}\tilde{R}_i^{-1}\tilde{B}'\tilde{P}_i\hat{\tilde{x}} + T^{-1}[(k+1)\tilde{B}'\tilde{R}^{-1}\tilde{P}_i\Pi + \Gamma]w, \\ \quad \text{for } \hat{\tilde{x}} \in \varepsilon_i \setminus \varepsilon_{i+1}, i = 0, 1, \dots, N-1 \\ u_N = v - (k+1)T^{-1}\tilde{R}_N^{-1}\tilde{B}'\tilde{P}_N\hat{\tilde{x}} + T^{-1}[(k+1)\tilde{B}'\tilde{R}_N^{-1}\tilde{P}_N\Pi + \Gamma]w \quad \text{for } \hat{\tilde{x}} \in \varepsilon_N \end{cases} \quad (30)$$

where $\hat{\tilde{x}} = [\hat{x}', v']'$.

We next proceed to show that,

Point 1. there exists a $k_1^* > 0$, and for each $k \geq k_1^*$, there exists an $\ell_1^*(k) > 0$ such that, for all $\ell \geq \ell_1^*(k)$, $k \geq k_1^*$, Item 1 of Problem 2 holds;

Point 2. there exists a $k_2^* > 0$, and for each $k \geq k_2^*$, there exists an $\ell_2^*(k) > 0$ such that, for all $\ell \geq \ell_2^*(k)$, $k \geq k_2^*$, Item 2 of Problem 2 holds,

from which the results of Theorem 5.2 follows with $k^* = \max\{k_1^*, k_2^*\}$, and for each $k \geq k^*$, $\ell^*(k) = \max\{\ell_1^*(k), \ell_2^*(k)\}$.

Proof of Point 1.

To show the existence of k_1^* , and $\ell_1^*(k)$ for each $k \geq k_1^*$, let us write out the closed-loop system,

$$\begin{cases} \dot{\tilde{x}} = \tilde{A}\tilde{x} + \tilde{B}\text{sat}_\Delta(-(k+1)\tilde{R}_i^{-1}\tilde{B}'\tilde{P}_i\tilde{x} + (k+1)\tilde{R}_i^{-1}\tilde{B}'\tilde{P}_i\tilde{\phi} + d), \\ \quad \hat{\tilde{x}} \in \varepsilon_i \setminus \varepsilon_{i+1}, \tilde{x}(0) \in \mathcal{X} \times \mathcal{V} \\ \dot{\phi} = (A - L(\ell)C)\phi, \phi(0) \in \mathcal{X} \end{cases} \quad (31)$$

where $\tilde{x} = [x', v']'$, $\tilde{\phi} = [\phi', 0]' = [x' - \hat{x}', 0]'$. Noting that $A - L(\ell)C$ is Hurwitz for all $\ell > 0$ and hence $\lim_{t \rightarrow \infty} \phi(t) = 0$, we need only to show that, for any $k \geq k_1^*$ and $\ell \geq \ell_1^*(k)$, $(\hat{x}(t), v(t))$ enters $\mathcal{X}_\infty \times \mathcal{V}_\infty$ in a finite time and remains in it thereafter, where \mathcal{X}_∞ and \mathcal{V}_∞ both contain the origin of as an interior point and are such that $\mathcal{X}_\infty \times \mathcal{V}_\infty \times \mathcal{X}_\infty$ is strictly inside the set \mathcal{Z}_∞ . To do this, we recall that $\mathcal{X} \times \mathcal{V}$ is strictly inside the set Ξ . Since the dynamics of \tilde{x} is a linear system

driven by a bounded input whose bound is independent of ℓ , it follows from (28) that there exist an $\ell_{1a}^* > 0$ and a $T_0 \geq 0$ such that for all $\ell \geq \ell_{1a}^*$,

$$\tilde{x}(T_0) = (x(T_0), v(T_0)) \in \Xi, \hat{\tilde{x}}(T_0) = (\hat{x}(T_0), v(T_0)) \in \Xi, \forall (x(0), v(0)) \in \mathcal{X} \times \mathcal{V}$$

Hence $(\hat{x}(T_0), v(T_0)) \in \varepsilon_0$. We next show that there exists a $k_1^* > 0$ and for each $k \geq k_1^*$ there exists an $\ell_1^*(k) > 0$ such that, for any $\ell \geq \ell_1^*(k)$, $k \geq k_1^*$, the trajectory $(\hat{x}(t), v(t))$ will enter $\mathcal{X}_\infty \times \mathcal{V}_\infty$ in a finite time and remain in it thereafter. This can be done in two steps. In the first step, we show that, for each $i = 0$ to $N - 1$, there exists a k_{1i}^* and for each $k \geq k_{1i}^*$ there exists an $\ell_{1i}^*(k) \geq 0$ such that for all $\ell \geq \ell_{1i}^*(k)$, $k \geq k_{1i}^*$, all trajectories $(\hat{x}(t), v(t))$ starting from $\varepsilon_i \setminus \varepsilon_{i+1}$ will remain in ε_i and enter into the inner level set ε_{i+1} in a finite time. This in turn implies that, for any $\ell \geq \max\{\ell_{10}^*(k), \ell_{11}^*(k), \dots, \ell_{1N-1}^*(k)\}$, $k \geq \max\{k_{10}^*, k_{11}^*, \dots, k_{1N-1}^*\}$, all trajectories $(\hat{x}(t), v(t))$ starting from ε_0 will enter the inner-most level set ε_N in a finite time. The second step of the proof is to show that there exists a $k_{1N}^* > 0$ and for each $k \geq k_{1N}^*$ there is an $\ell_{1N}^* \geq \ell_{1a}^*(k)$ such that, for all $\ell \geq \ell_{1N}^*(k)$, $k \geq k_{1N}^*$, all the trajectories $(\hat{x}(t), v(t))$ starting from ε_N will enter and remain in the set $\mathcal{X}_\infty \times \mathcal{V}_\infty$ in a finite time. Once these two steps are completed, the proof of Point 1 is then completed by taking $k_1^* = \max\{k_{10}^*, k_{11}^*, \dots, k_{1N}^*\}$ and for each $k \geq k_1^*$, $\ell_1^*(k) = \max\{\ell_{10}^*(k), \ell_{11}^*(k), \dots, \ell_{1N}^*(k)\}$.

We start by considering the closed-loop system (31) for $(\hat{x}, v) \in \varepsilon_i \setminus \varepsilon_{i+1}$ and $t \geq T_0$,

$$\begin{cases} \dot{\tilde{x}} = \tilde{A}\tilde{x} + \tilde{B}\text{sat}_\Delta(-(k+1)\tilde{R}_i^{-1}\tilde{B}'\tilde{P}_i\tilde{x} + (k+1)\tilde{R}_i^{-1}\tilde{B}'\tilde{P}_i\tilde{\phi} + d), \\ \dot{\tilde{\phi}} = (A - L(\ell)C)\tilde{\phi}, \tilde{\phi}(0) \in \mathcal{X} \end{cases} \quad (32)$$

Pick the Lyapunov function

$$V_i(\hat{\tilde{x}}) = \hat{\tilde{x}}' \tilde{P}_i \hat{\tilde{x}}. \quad (33)$$

Recalling that

$$\hat{\tilde{x}} = \begin{bmatrix} \hat{x} \\ v \end{bmatrix} = \begin{bmatrix} x - \phi \\ v \end{bmatrix} = \tilde{x} - \tilde{\phi}$$

we have

$$V_i(\hat{\tilde{x}}) = \tilde{x}' \tilde{P}_i \tilde{x} - 2\tilde{x}' \tilde{P}_i \tilde{\phi} + \tilde{\phi}' \tilde{P}_i \tilde{\phi} \quad (34)$$

Now for any $k > 0$, let $\ell_{1b}^*(k) \geq \ell_{1a}^*$ be such that, for any $\ell \geq \ell_{1b}^*(k)$, $k > 0$, and for all $(x(0), v(0), \hat{x}(0)) \in \mathcal{X} \times \mathcal{V} \times \mathcal{X}$,

$$\begin{cases} \|(k+1)\tilde{R}_i^{-1}\tilde{B}'\tilde{P}_i\tilde{\phi}(t)\| \leq D, \quad \forall t \geq T_0 \\ \|\tilde{x}'\tilde{P}_i[A - L(\varepsilon)C]\tilde{\phi}(t)\| \leq \frac{D^2}{k}, \quad \forall(\hat{x}, v) \in \varepsilon_0, \quad \forall t \geq T_0 \\ \|\tilde{\phi}'(t)\tilde{P}_i[\tilde{A}\tilde{x} + \tilde{B}\text{sat}_\Delta(\cdot)]\| \leq \frac{D^2}{k}, \quad \forall(\hat{x}, v) \in \varepsilon_0, \quad \forall t \geq T_0 \\ \|\tilde{\phi}'(t)\tilde{P}_i[A - L(c)C]\tilde{\phi}(t)\| \leq \frac{D^2}{k}, \quad \forall t \geq T_0 \end{cases} \quad (35)$$

Let $\tilde{d} = (k+1)\tilde{R}_i^{-1}\tilde{B}'\tilde{P}_i\tilde{\phi}(t) + d$. In view of Assumption 2.3 and (35), we have that

$$\|\tilde{d}\| \leq 2D, \quad \forall t \geq T_0 \quad (36)$$

Viewing \tilde{d} as disturbance and in view of (35), it follows from the same arguments as used for arriving at (21) that

$$\dot{V}_i \leq -\tilde{x}'\tilde{x} + \frac{22D^2}{k} \leq -\frac{1}{\lambda_{\max}(\tilde{P}_{i+1})} \left(\tilde{x}'\tilde{P}_{i+1}\tilde{x} - \frac{22\lambda_{\max}(\tilde{P}_{i+1})D^2}{k} \right), \quad \forall \hat{x} \in \varepsilon_i \setminus \varepsilon_{i+1} \quad (37)$$

For each k , let $\ell_{1i}^*(k) \geq \ell_{1b}^*(k)$ be such that, for all $t \geq T_0$, $\hat{x}'\tilde{P}_{i+1}\hat{x} \geq \rho_{i+1}$ implies that $\tilde{x}'\tilde{P}_{i+1}\tilde{x} \geq \frac{1}{2}\rho_{i+1}$. Also let

$$k_{1i}^*(D) = \frac{45\lambda_{\max}(\tilde{P}_{i+1})D^2}{\rho_{i+1}} \quad (38)$$

We then have that, for all $\ell \geq \ell_{1i}^*(k)$, $k \geq k_{1i}^*$,

$$\dot{V}_i < 0, \quad \forall \hat{x} \in \varepsilon_i \setminus \varepsilon_{i+1} \quad (39)$$

which implies that all trajectories $(\hat{x}(t), v(t))$ of the closed-loop system starting from $\varepsilon_i \setminus \varepsilon_{i+1}$ will remain in ε_i and enter the inner set ε_{i+1} in a finite time and remain in it thereafter.

For the second step of the proof of Point 1, let $\rho_{N+1} \in (0, \rho_N]$ be such that $\varepsilon(\tilde{P}_N, \rho_{N+1}) \subset \mathcal{X}_\infty \times \mathcal{V}_\infty$. The existence of such a ρ_{N+1} is due to the fact that $\mathcal{X}_\infty \times \mathcal{V}_\infty$ contains the origin of the state space as an interior point. We will have,

$$\dot{V}_N \leq -\frac{1}{\lambda_{\max}(\tilde{P}_N)} \left(\tilde{x}'\tilde{P}_N\tilde{x} - \frac{22\lambda_{\max}(\tilde{P}_N)D^2}{k} \right), \quad \forall \hat{x} \in \varepsilon_N \quad (40)$$

For each $k > 0$, let $\ell_{iN}^*(k) \geq \ell_{1b}^*(k)$ be such that, for all $t \geq T_0$, $\hat{x}'\tilde{P}_N\hat{x} \geq \rho_{N+1}$ implies that $\tilde{x}'\tilde{P}_N\tilde{x} \geq \frac{1}{2}\rho_{N+1}$. Also let

$$k_{iN}^*(D) = \frac{45\lambda_{\max}(\tilde{P}_N)D^2}{\rho_{N+1}} \quad (41)$$

We then have that, for all $\ell \geq \ell_{1N}^*(k)$, $k \geq k_{iN}^*$,

$$\dot{V}_N < 0, \quad \forall \hat{x} \in \varepsilon_N \setminus \varepsilon(\tilde{P}_N, \rho_{N+1}) \quad (42)$$

which shows that all trajectories $(\hat{x}(t), v(t))$ of the closed-loop system that start from ε_N will remain in ε_N and enter the set $\varepsilon(\tilde{P}_N, \rho_{N+1}) \subset \mathcal{X}_\infty \times \mathcal{V}_\infty$ in a finite time and remain in it thereafter.

This completes the proof of Point 1. We next proceed to prove Point 2.

Proof of Point 2.

Let us start with the introduction of an invertible, triangular coordinate change $\xi = \tilde{x} - \Pi w$.

Using Condition 1 of the theorem, we can write the closed-loop system as follows,

$$\begin{cases} \dot{\xi} = \tilde{A}\xi + \tilde{B}[\text{sat}_\Delta(-(k+1)\tilde{R}_i^{-1}\tilde{B}'\tilde{P}_i\xi + (k+1)\tilde{R}_i^{-1}\tilde{B}'\tilde{P}_i\tilde{\phi} + \Gamma w + d) - \Gamma w], \\ \quad \hat{x} \in \varepsilon_i \setminus \varepsilon_{i+1}, \quad \tilde{x}(0) \in \mathcal{X} \times \mathcal{V} \\ \dot{\phi} = (A - L(\ell)C)\phi, \quad \phi(0) \in \mathcal{X} \end{cases} \quad (43)$$

which, except for the terms Γw , is exactly the same as (31). The rest of the proof follows the arguments used in the proof of Point 2 of Theorem 4.1 and those used in the proof of Point 1 of the current theorem, Theorem 5.2. \square

6 Design for Robust Tracking for an F-16 Aircraft

In this section, the applicability and effectiveness of the proposed combined PLC/LHG design algorithm is demonstrated with an F-16 fighter aircraft derivative. At the flight condition corresponding to an altitude of 10,000 feet and a Mach number of 0.7, the second order pitch plane dynamics (short period) of this aircraft is given by [16],

$$\begin{cases} \begin{bmatrix} \dot{\alpha} \\ \dot{q} \end{bmatrix} = \begin{bmatrix} -1.1500 & 0.9937 \\ 3.7240 & -1.2600 \end{bmatrix} \begin{bmatrix} \alpha \\ q \end{bmatrix} + \begin{bmatrix} -0.1770 \\ -19.5000 \end{bmatrix} \delta, \quad |\alpha(0)| \leq 0.1\text{rad}, |q(0)| \leq 0.5\text{rad/sec} \\ \dot{\delta} = \text{sat}_1(-T\delta + Tu + d), \quad |\delta(0)| \leq 0.1\text{rad} \\ \dot{w} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} w, \quad |w_i(0)| \leq 0.5\text{rad/sec}, i = 1, 2 \\ e = [0 \quad 1] \begin{bmatrix} \alpha \\ q \end{bmatrix} + [0 \quad 1] w \\ y = [1 \quad 0] \begin{bmatrix} \alpha \\ q \end{bmatrix} \end{cases}$$

where α , q and δ are respectively the trim value of the angle of attack, pitch rate and stabilizer deflection. Here the actuator is rate limited to ± 1.0 rad/sec. The constants $T = 20$ sec, which corresponds to an actuator bandwidth of 20 rad/sec.

It is straightforward to see that the above system is in the form of (1). Our goal here is to design an effective robust tracking control law that will cause the pitch rate q to track a class of sinusoidal reference signals in the presence of actuator disturbance d . The control law to be designed is the combined PLC/LHG feedback law proposed in the previous two sections.

State Feedback Design

In the case that both states α and q are available for feedback, we will design state feedback laws as proposed in Section 4. We begin by checking the conditions of Theorem 4.1. It is straightforward to check that Condition 1 is satisfied. More specifically, the matrices

$$\Pi = \begin{bmatrix} -0.4141 & -0.4993 \\ 0 & -1 \\ -0.1304 & -0.0307 \end{bmatrix}, \quad \Gamma = [0.0307 \quad -0.1304] \quad (44)$$

solve the linear matrix equations (13). It is also straightforward to verify that $|\Gamma w(t)| \leq 0.2$ for all $w(0) \in \mathcal{W}$ and for all $t \geq 0$. Hence, Condition 2 of the theorem is satisfied with $\delta = 0.8$.

Following the design procedure given in Section 4, we get

$$\Xi = \{(\xi_1, \xi_2, \xi_3) : |\xi_1| \leq 0.5567, |\xi_2| \leq 1, |\xi_3| \leq 0.1805\}$$

and find,

$$\rho_0 = 9.1920, \quad \epsilon_0 = 26.8917, \quad P_0 = \begin{bmatrix} 2.0409 & 1.0217 & -10.2609 \\ 1.0217 & 0.7091 & -5.7828 \\ -10.2609 & -5.7828 & 78.6731 \end{bmatrix}$$

Choosing $N = 5$ and $\Delta\rho = 0.5$, we obtain a feedback law of the form (17) with,

$$\rho_1 = 4.5960, \quad \epsilon_1 = 15.0397, \quad P_1 = \begin{bmatrix} 1.5809 & 0.7792 & -6.9023 \\ 0.7792 & 0.5805 & -4.0412 \\ -6.9023 & -4.0412 & 49.2153 \end{bmatrix}$$

$$\rho_2 = 2.2980, \quad \epsilon_2 = 8.4670, \quad P_2 = \begin{bmatrix} 1.2568 & 0.6043 & -4.7011 \\ 0.6043 & 0.4849 & -2.8789 \\ -4.7011 & -2.8789 & 31.1964 \end{bmatrix}$$

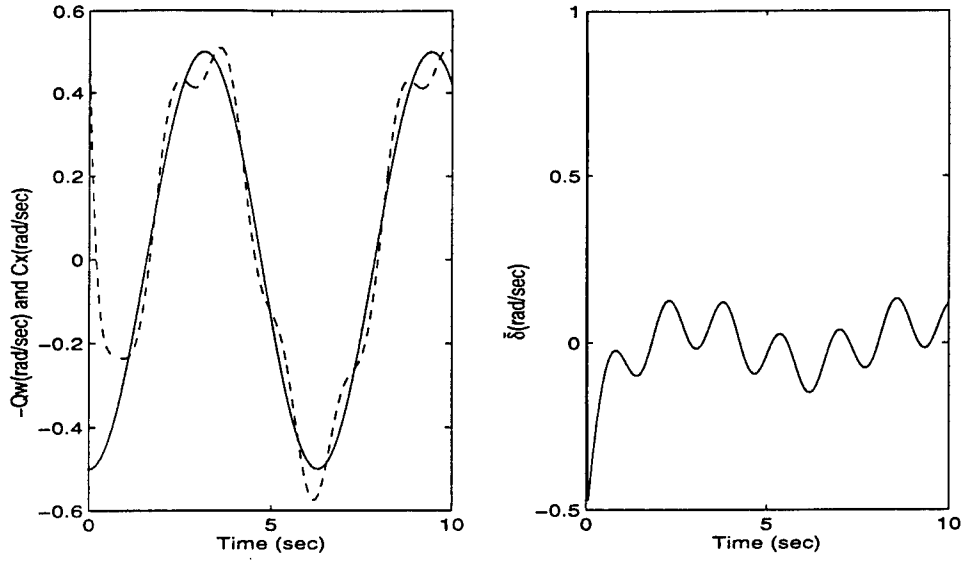


Figure 1: Combined PLC/LHG design ($N = 5$, $k = 40$). $(\alpha(0), q(0), \delta(0), w_1(0), w_2(0)) = (-0.1, 0.5, 0.1, 0, 0.5)$ and $d = 5 \sin(4t + 3)$. The left plot: the plant output Cx (dashed line) and the reference signal $-Qw$ (solid line). The right plot: the actuator rate $\dot{\delta}$.

$$\rho_3 = 1.1490, \quad \epsilon_3 = 4.7942, \quad P_3 = \begin{bmatrix} 1.0252 & 0.4753 & -3.2344 \\ 0.4753 & 0.4118 & -2.0852 \\ -3.2344 & -2.0852 & 20.0040 \end{bmatrix}$$

$$\rho_4 = 0.5745, \quad \epsilon_4 = 2.7281, \quad P_4 = \begin{bmatrix} 0.8580 & 0.3784 & -2.2435 \\ 0.3784 & 0.3541 & -1.5314 \\ -2.2435 & -1.5314 & 12.9550 \end{bmatrix}$$

$$\rho_5 = 0.2873, \quad \epsilon_5 = 1.5590, \quad P_5 = \begin{bmatrix} 0.7363 & 0.3045 & -1.5666 \\ 0.3045 & 0.3075 & -1.1376 \\ -1.5666 & -1.1376 & 8.4611 \end{bmatrix}$$

We note here again that this control law reduces to an LHG feedback control law if $N = 0$ and to a PLC law if $k = 0$.

Figs. 1 and 2 are simulation results. In these simulations, $(\alpha(0), q(0), \delta(0), w_1(0), w_2(0)) = (-0.1, 0.5, 0.1, 0, 0.5)$, and $d = 5 \sin(4t + 3)$. They show good utilization of the available actuator rate capacity, and that the degree of tracking accuracy in the presence of actuator disturbance increases as the value of the parameter k increases.

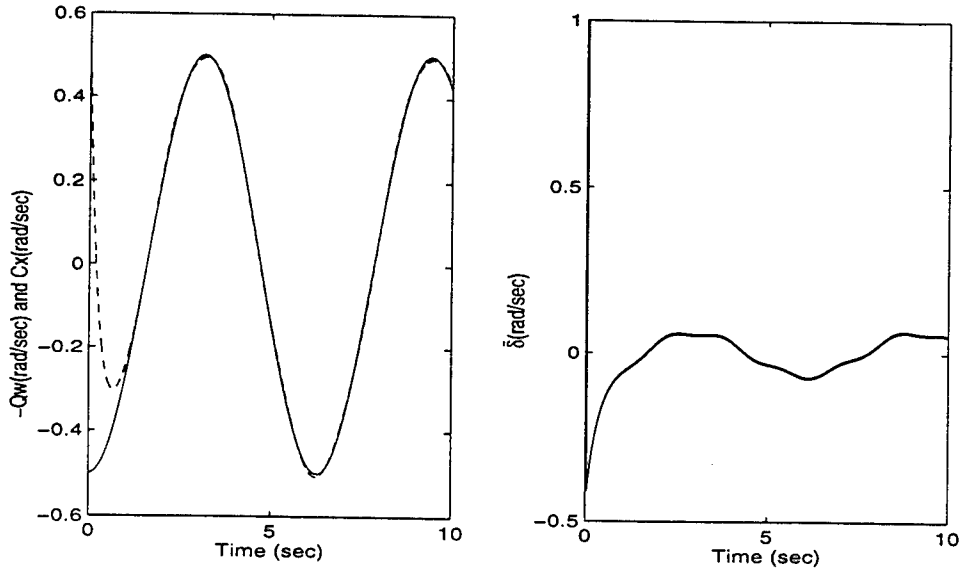


Figure 2: Combined PLC/LHG design ($N = 5$, $k = 400$). $(\alpha(0), q(0), \delta(0), w_1(0), w_2(0)) = (-0.1, 0.5, 0.1, 0, 0.5)$ and $d = 5 \sin(4t + 3)$. The left plot: the plant output Cx (dashed line) and the reference signal $-Qw$ (solid line). The right plot: the actuator rate $\dot{\delta}$.

We next compare the performance of the combined PLC/LHG design with both the PLC and the LHG design. To compare with the PLC laws, we set $k = 0$ in (17). The simulation (Fig. 3) shows that in the presence of the same disturbance, the aircraft becomes unstable. To compare with the LHG feedback law, we set $N = 0$ in (17). The simulation (Fig. 4) shows that the combined PLC/LHG feedback law results in much better transient performance. By increasing the value of N , the transience performance can be further improved.

Output Feedback Design

In the case that only the output (i.e., α) is available for feedback, we design output feedback laws as proposed in Section 5. We have already checked in the state feedback case that all condition of Theorem 5.2 are satisfied except Assumption 5.1, which is also satisfied. Following Step 1 of the output feedback design algorithm, we obtain the fast observer as given by (27) with

$$L(\ell) = [2\ell - 2.41 \quad 1.0063\ell^2 - 2.536\ell + 5.3216]'$$

We note here that, with this choice of $L(\ell)$,

$$\lambda(A - L(\ell)C) = \{-\ell, -\ell\}$$

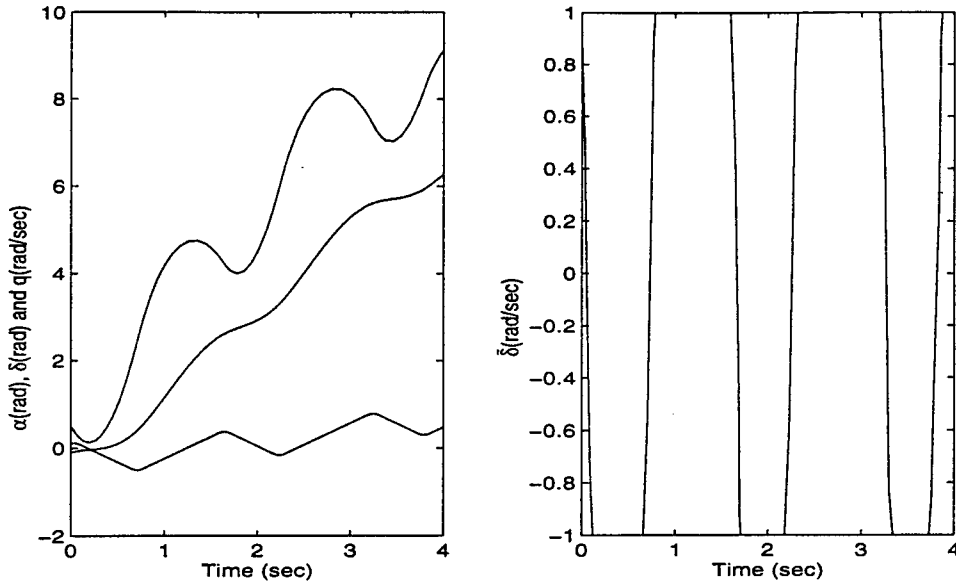


Figure 3: PLC Design ($N = 5, k = 0$). $(\alpha(0), q(0), \delta(0), w_1(0), w_2(0)) = (-0.1, 0.5, 0.1, 0, 0.5)$ and $d = 5 \sin(4t + 3)$. The left plot: states of the plant and the actuator. The right plot: the actuator rate $\dot{\delta}$.

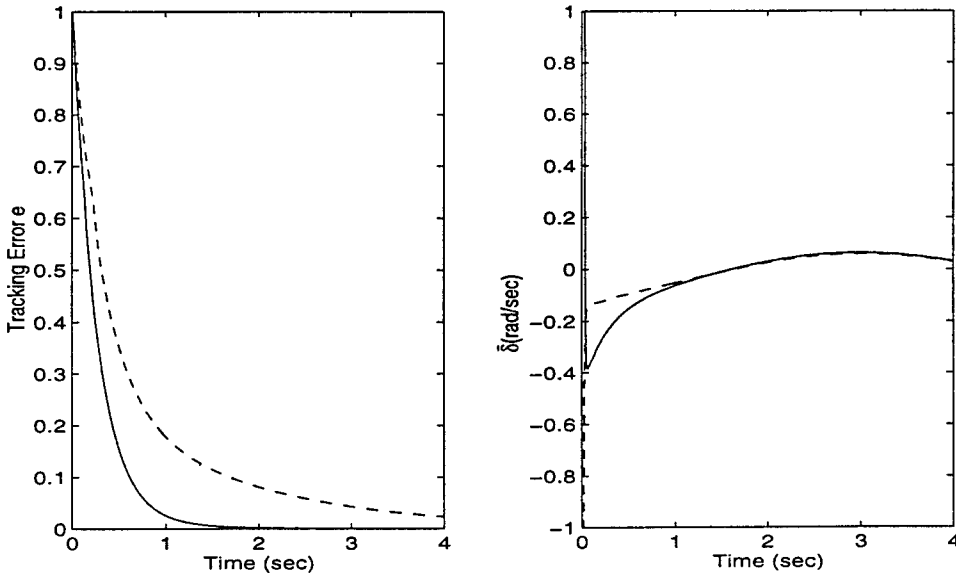


Figure 4: Transience performance: PLC/LHG ($N = 5, k = 40$, solid lines) vs. LHG ($N = 0, k = 40$, dashed lines). $(\alpha(0), q(0), \delta(0), w_1(0), w_2(0)) = (-0.1, 0.5, 0.1, 0, 0.5)$ and $d = 0$. The left plot: tracking error e . The right plot: the actuator rate $\dot{\delta}$.

Proceeding with Step 2 of the design, we first choose

$$\Xi = \{(\xi_1, \xi_2, \xi_3) : |\xi_1| \leq 0.56, |\xi_2| \leq 1.01, |\xi_3| \leq 0.185\}$$

and find,

$$\rho_0 = 9.8822, \quad \epsilon_0 = 28.5865, \quad P_0 = \begin{bmatrix} 2.0993 & 1.0521 & -10.7036 \\ 1.0521 & 0.7250 & -6.0107 \\ -10.7036 & -6.0107 & 82.6928 \end{bmatrix}$$

Choosing $N = 5$ and $\Delta\rho = 0.5$, we obtain a state feedback law of the form (29) with,

$$\rho_1 = 4.9411, \quad \epsilon_1 = 15.9758, \quad P_1 = \begin{bmatrix} 1.6217 & 0.8010 & -7.1897 \\ 0.8010 & 0.5921 & -4.1914 \\ -7.1897 & -4.1914 & 51.6538 \end{bmatrix}$$

$$\rho_2 = 2.4705, \quad \epsilon_2 = 8.9881, \quad P_2 = \begin{bmatrix} 1.2857 & 0.6201 & -4.8911 \\ 0.6201 & 0.4937 & -2.9803 \\ -4.8911 & -2.9803 & 32.6997 \end{bmatrix}$$

$$\rho_3 = 1.2353, \quad \epsilon_3 = 5.0865, \quad P_3 = \begin{bmatrix} 1.0460 & 0.4871 & -3.3619 \\ 0.4871 & 0.4186 & -2.1552 \\ -3.3619 & -2.1552 & 20.9446 \end{bmatrix}$$

$$\rho_4 = 0.6176, \quad \epsilon_4 = 2.8930, \quad P_4 = \begin{bmatrix} 0.8731 & 0.3873 & -2.3302 \\ 0.3873 & 0.3596 & -1.5807 \\ -2.3302 & -1.5807 & 13.5511 \end{bmatrix}$$

$$\rho_5 = 0.3088, \quad \epsilon_5 = 1.6526, \quad P_5 = \begin{bmatrix} 0.7473 & 0.3113 & -1.6261 \\ 0.3113 & 0.3120 & -1.1730 \\ -1.6261 & -1.1730 & 8.8436 \end{bmatrix}$$

Implementing the state feedback law with the fast observer state, both obtained as above, we have the output feedback law in the form of (30).

Some simulation results are shown in Fig. 5. In the simulation, $(\alpha(0), q(0), \delta(0), w_1(0), w_2(0)) = (-0.1, 0.5, 0.1, 0, 0.5)$, $d(t) = 5 \sin(4t + 3)$, which are the same as in the simulations in the state feedback case, and $(\hat{x}_1(0), \hat{x}_2(0)) = (0, 0)$. With $k = 400$, $\ell = 1000$, these simulations results are almost the same as those shown in Fig. 2.

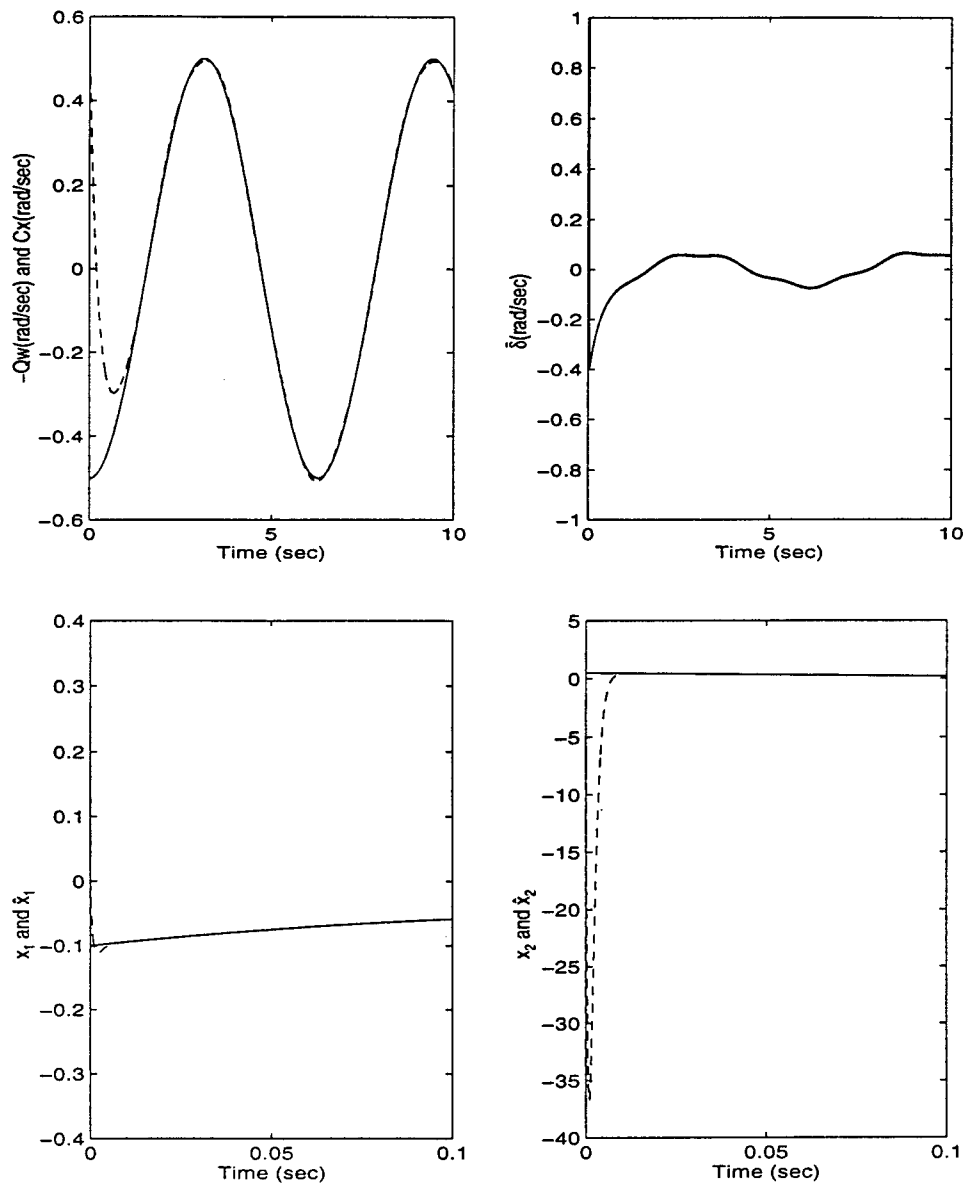


Figure 5: PLC Design ($N = 5$, $k = 0$). $(\alpha(0), q(0), \delta(0), w_1(0), w_2(0), \hat{x}_1(0), \hat{x}_2(0)) = (-0.1, 0.5, 0.1, 0, 0.5, 0, 0)$ and $d = 5 \sin(4t + 3)$. The upper left plot: the plant output Cx (dashed line) and the reference signal $-Qw$ (solid line). The upper right plot: the actuator rate $\dot{\delta}$. The lower plots: plant states $x(t)$ (solid line) and their estimates \hat{x} (dashed line).

7 Conclusions

A design technique we recently proposed for stabilization of a linear system with rate-limited actuators is utilized to design feedback laws that cause the system output to track a desired command signal. This design method is based on two design techniques recently developed for linear systems with position limited actuators and retains the advantages of both design techniques, while avoiding their disadvantages. Application of the proposed design to an F-16 fighter aircraft demonstrates the applicability, robustness and effectiveness of this novel control law.

References

- [1] J.M. Berg, K.D. Hammett, C.A. Schwartz, and S.S. Banda, "An analysis of the destabilizing effect of daisy chained rate-limited actuators," *IEEE Trans. Control Sys. Tech.*, Vol. 4, No. 2, 1996.
- [2] D.S. Bernstein and A.N. Michel, "A chronological bibliography on saturating actuators," *International Journal of Robust and Nonlinear Control*, Vol.5, No.5, pp.375-380, 1995.
- [3] M.A. Dornheim, "Report pinpoints factors leading to YF-22 crash," *Aviation Week Space Technol.*, pp. 53-54, Nov. 9, 1992.
- [4] B. Francis, "The linear multivariable regulator problem," *SIAM J. Control Optimization*, Vol. 15, 486-505, 1984.
- [5] J.M. Lenorovitz, "Gripen control problems resolved through in-flight, ground simulations," *Aviation Week Space Technol.*, pp. 74-75, June 18, 1990.
- [6] Z. Lin, "Semi-global stabilization of linear systems with position and rate limited actuators," *Systems & Control Letters*, Vol. 30, No. 1, pp. 1-11, 1997.
- [7] Z. Lin, "Global control of linear systems with saturating actuators," *Proc. of the 35th CDC*, pp. 4357-4362, 1996.
- [8] Z. Lin, M. Pachter and S. Banda, "Toward improvement on tracking performance - non-linear feedback for linear systems," *International Journal of Control*, to appear. Also in the *Proceedings of the 2nd Asian Control Conference*, Vol. II, pp. 527-530, 1997.

- [9] Z. Lin, M. Pachter, S. Banda and Y. Shamash, "Stabilizing feedback design for linear systems with rate limited actuators," *Control of Uncertain Systems with Bounded Inputs*, eds. S. Tarbouriech and Germain Garcia, Lecture Notes in Control and Information Sciences, Springer-Verlag, Vol. 227, pp. 173-186, 1997.
- [10] Z. Lin, M. Pachter, S. Banda, and Y. Shamash, "Feedback design for robust tracking of linear with rate limited actuators," *Proceedings of 1997 AIAA Guidance, Navigation and Control Conference*, AIAA paper 97-3609, pp.780-788, 1997.
- [11] Z. Lin, M. Pachter, S. Banda, and Y. Shamash, "Feedback design for robust tracking of linear with rate limited actuators (with output feedback)," submitted for publication.
- [12] Z. Lin, A. Saberi and B.M. Chen, *Linear Systems Toolbox*, (Commercially available through *A.J. Controls Inc.* Seattle, Washington.) Washington State University Report No. EE/CS 0097, 1991.
- [13] Z. Lin, A. Stoorvogel and A. Saberi, "Output regulation for linear systems subject to input saturation," *Automatica*, Vol. 32, No 1, pp.29-47, 1996.
- [14] A. Megretski, "Output feedback stabilization with saturated control: making the input-output map L_2 -bounded," preprint. See also " L_2 BIBO output feedback stabilization with saturated control," *Proc. 13th Triennial World Congress of IFAC*, vol. D, pp. 435-440, 1996.
- [15] P. Miotto, J.M. Shewchun, E. Feron, and J.D. Paduano, "High performance bounded control synthesis with application to the F18 HARV. *Proc. of the AIAA GNC*, San Diego, CA, July 1996.
- [16] M. Pachter, P.R. Chandler and M. Mears, "Reconfigurable tracking control with saturation," *AIAA Journal of Guidance, Control and Dynamics*, Vol. 18, pp 1016-1022, 1995.
- [17] C.A. Shifrin, "Gripen likely to fly again soon," *Aviation Week Space Technol.*, pp. 72, Aug. 23, 1993.
- [18] A. Saberi, Z. Lin and A.R. Teel, "Control of linear systems subject to input saturation," *IEEE Trans. Auto. Contr.*, Vol. 41, No.3, pp. 368-378, 1996.

- [19] P. Sannuti and A. Saberi, "A special coordinate basis of multivariable linear systems – finite and infinite zero structure, squaring down and decoupling," *International J. Control*, Vol. 45, pp. 1655-1704, 1987.
- [20] G.F. Wredenhagen and P.R. Belanger, "Piecewise-linear LQ control for systems with input constraints," *Automatica*, vol. 30, pp. 403-416, 1994.

EXPERIMENTAL AND COMPUTATIONAL INVESTIGATIONS OF BROMINE AND
IODINE CHEMISTRY IN FLAME SUPPRESSION

Paul Marshall
Associate Professor
Department of Chemistry

University of North Texas
PO Box 305070, Denton, Texas 76203-5070

Final Report for:
Summer Faculty Extension Program
Wright Laboratory, Materials Directorate, Wright-Patterson AFB, Ohio 45433

Sponsored by:
Air Force Office of Scientific Research
Bolling AFB, Washington, D.C.

and

Wright Laboratory, Wright-Patterson AFB

December 1997

EXPERIMENTAL AND COMPUTATIONAL INVESTIGATIONS OF BROMINE AND IODINE CHEMISTRY IN FLAME SUPPRESSION

Paul Marshall

Associate Professor

Department of Chemistry

University of North Texas

PO Box 305070, Denton, Texas 76203-5070

Abstract

The kinetics of the reaction of hydroxyl radicals with trifluoroiodomethane were investigated by the flash photolysis - resonance fluorescence technique. A rate constant of $k = 9.9 \times 10^{-12} \exp(-12.7 \text{ kJ mol}^{-1}/RT) \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$ was measured over the temperature range 280-450 K with accuracy limits of 10% (450 K) to 30% (280 K). Different product channels were investigated by ab initio methods, and the dominant products are $\text{CF}_3 + \text{HOI}$. The enthalpy of formation of hypoiodous acid was analyzed with Gaussian-2 theory, in conjunction with G2 energies for INO , ICN , ClCN and other species. The transition state and reaction coordinate for $\text{OH} + \text{CF}_3\text{I}$ was characterized at the G2(MP2) level, and the results suggest a negligible barrier to the reverse reaction of $\text{CF}_3 + \text{HOI}$, so that the measured forward activation energy can be used to derive $\Delta_f H_{298}(\text{HOI}) = -68 \pm 5 \text{ kJ mol}^{-1}$. The implications of the kinetics and thermochemistry for iodine chemistry in flames and the atmosphere are discussed, and for the range 280 - 2000 K a proposed rate expression is $k = 2.9 \times 10^{-16} (T/\text{K})^{1.5} \exp(-960 \text{ K}/T) \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$.

EXPERIMENTAL AND COMPUTATIONAL INVESTIGATIONS OF BROMINE AND IODINE CHEMISTRY IN FLAME SUPPRESSION

Paul Marshall

1. Introduction

Trifluoroiodomethane is a possible replacement for Halon-1301 (CF_3Br) as a fire suppression agent in some applications,¹ following the ban on Halon production under the Montreal Protocol on Substances that Deplete the Ozone Layer. A potentially important step in the combustion chemistry of CF_3I is the reaction



whose kinetics have been investigated previously only at room temperature.^{2,3} The two earlier studies yielded rate constants k_1 that differ by a factor of four. One aim of the present work is to determine the Arrhenius parameters of the isolated elementary reaction for the first time, in order to permit estimation of k_1 at flame temperatures. The second aim is to measure k_1 at below room temperature, in order to assess the atmospheric lifetime of CF_3I at the average tropospheric temperature of around 280 K. The third aim is to assess the main products of the reaction. Finally, the kinetic measurements are combined with ab initio information about the potential energy surface to obtain the thermochemistry of a dominant product, hypoiodous acid. HOI is a major product of the reaction of ground state atomic oxygen with $\text{C}_2\text{H}_5\text{I}$ and larger iodoalkanes,^{4,5} and recently has been studied by photoionization⁶ and FT-IR emission techniques.⁷ $\Delta_f H_{298}(\text{HOI})$ has not previously been determined experimentally, despite its importance to understanding the photodissociation dynamics⁸ and photoionization⁶ of hypohalous acids and the roles of HOI in the mechanisms of iodine-mediated flame suppression⁹ and stratospheric iodine chemistry.¹⁰ Here we employ an ab initio potential energy surface (PES) to estimate the activation energy for the reverse rate constant k_{-1} , and thus obtain the thermochemistry of reaction 1. The result is checked against results from Gaussian-2 (G2) theory, which for first and second-row elements has a typical accuracy of about 8 kJ mol⁻¹.¹¹ As will be shown, our measured and computed $\Delta_f H_{298}(\text{HOI})$ are in good accord but lie around 15 -20 kJ mol⁻¹ below some recently proposed values, and thus HOI is more stable than previously thought.

2. Experimental method

The gas-handling system and the flash photolysis - resonance fluorescence reactor are described elsewhere.^{12,13,14} Briefly, OH radicals were generated by flash lamp photolysis of water vapor through magnesium fluoride optics ($\lambda > 120$ nm), in the presence of excess CF_3I in a buffer of Ar gas. $[\text{OH}]$ was monitored as a function of time by resonance fluorescence at $\lambda \approx 307$ nm, excited from a flow of 2% H_2O vapor in Ar through a microwave-powered discharge lamp, and isolated with a band pass filter before detection by a photomultiplier tube with photon-counting electronics. The signals I_t from typically 100-500 pulses were averaged before analysis by non-linear least-squares fitting to an exponential decay plus a steady background B from scattered resonance radiation:¹⁵

$$I_t = A \exp(-k_{\text{ps1}} t) + B \quad (2)$$

where the pseudo-first-order decay coefficient is

$$k_{\text{ps1}} = k_{1,\text{eff}}[\text{CF}_3\text{I}] + k_{\text{diff}} \quad (3)$$

k_{diff} describes the first-order loss of OH by diffusion to the walls of the reactor and was typically 50-200 s^{-1} . The effective second-order rate constant $k_{1,\text{eff}}$ was obtained from the slope of plots of typically four to six k_{ps1} values vs $[\text{CF}_3\text{I}]$ from 0 to $[\text{CF}_3\text{I}]_{\text{max}}$. The glass reactor was modified from that employed to investigate OH + silane reactions¹⁴ by the addition of heating tape to reach elevated temperatures, up to about 450 K, and by the use of an ice bath to lower the temperature to about 280 K. The gas temperature at the reactor center was measured at the beginning and end of each experimental run with a moveable chromel-alumel thermocouple, calibrated against melting ice and boiling water. At each temperature studied various experimental parameters were varied to see how they influenced $k_{1,\text{eff}}$. These included the total pressure P and average gas velocity, which determine the average gas residence time in the reactor before photolysis τ_{res} , and the concentration of H_2O and the flash lamp energy F, which together influence the initial radical concentrations created photolytically.

3. Experimental results

Table 1 summarizes 43 measurements of $k_{1,\text{eff}}$ at five temperatures. The value of $k_{1,\text{eff}}$ was found to be independent of τ_{res} , which indicates that thermal decomposition of CF_3I was negligible. $k_{1,\text{eff}}$ was also independent of $[\text{H}_2\text{O}]$ and thus $[\text{OH}]_{t=0}$, which implies that OH did not react

significantly with itself or with the reaction products. $k_{1,\text{eff}}$ did vary significantly and systematically with the flash energy F which probably reflects secondary chemistry of OH with photolysis fragments of CF_3I . The concentration of these fragments will be approximately proportional to F , so the primary reaction was isolated from interferences by linear extrapolation to zero flash energy, as illustrated in Fig. 1. The five resulting $k_1(T)$ values are plotted in Arrhenius form in Fig. 2 and may be summarized over $T = 280\text{--}450\text{ K}$ as

$$k_1 = (9.9 \pm 2.2) \times 10^{-12} \exp(-12.7 \pm 0.6 \text{ kJ mol}^{-1}/RT) \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1} \quad (4)$$

where the errors are $\pm 1\sigma$ statistical precisions derived from a weighted fit that takes into account errors in both k_1 and T . Consideration of the coupling between the errors in the Arrhenius parameters through the covariance leads to $\pm 1\sigma$ precision estimates for k_1 from 7% at the central temperature of about 350 K to 10% at the extremes. We allow for possible systematic errors of $\pm 1 \times 10^{-14} \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$ in the measurements or extrapolations to arrive at an overall confidence interval for k_1 of about $\pm 27\%$ at 280 K improving to $\pm 10\%$ at 450 K.

4. Computational approach

The general principles of quantitative ab initio molecular orbital theory have been outlined elsewhere.^{16,17} The calculations were carried out using the GAUSSIAN 92 and 94 codes^{18,19} on SGI Power Challenge, Cray C-916 and Cray YMP-8 computers. Standard atomic basis sets were employed, except for iodine where the basis sets described by Glukhovtsev et al.²⁰ were used. The geometries of the reagent and product molecules were first optimized at the HF/6-31G(d) level of theory. Scans of the PES at this level indicated no barrier to reaction 1 beyond the endothermicity. Next, the molecular structures were refined at the MP2=full/6-31G(d) level, which includes a partial correction for electron correlation effects applied to all electrons. The optimized geometries are illustrated in Fig. 3. At this level a barrier to reaction 1 was located, and its geometry is also shown in Fig. 3. Vibrational frequencies were calculated at each of these stationary points on the PES, and they verify that a true TS geometry was obtained, with a single complex frequency that corresponds to motion along the reaction coordinate. The ab initio HF and MP2 frequencies were scaled by 0.8929 and 0.9646,²¹ respectively, to approximate the fundamental frequencies (listed in Table 2). As can be seen from Table 2, there is good agreement with the observed frequencies of the reactants and

products.^{5,27}

Single-point high level calculations were then carried out at the stationary points to characterize their energy, and these data are listed in Table 3. These are frozen-core, all-electron calculations without effective core potentials.²⁰ The QCISD(T)/6-311G(d,p) and MP2/6-311+G(3df,2p) results, based on the extension of Gaussian-2 (G2) theory¹¹ to Br and I by Glukhovtsev et al.²⁰, were combined according to the G2(MP2) methodology,²² where

$$E[\text{G2(MP2)}] = E[\text{QCISD(T)/6-311G(d,p)}] + E[\text{MP2/6-311+G(3df,2p)}] - \\ E[\text{MP2/6-311G(d,p)}] + \text{HLC} + \text{ZPE} \quad (5)$$

An empirical higher-level correction is based on the number of alpha and beta valence electrons:²¹ $\text{HLC} = -0.00019 n_\alpha - 0.00481 n_\beta$ hartrees (1 hartree ≈ 2625 kJ mol⁻¹), and $E[\text{G2(MP2)}]$ approximates a complete QCISD(T)/6-311+G(3df,2p) calculation. The zero-point vibrational energies (ZPE) are based on MP2 results. Because MP2 frequencies for the TS are available but HF frequencies are not, our methodology for the PES is a combination of the G2(MP2)²² and G2(ZPE=MP2)²¹ methods. It is noted that calculation of relative energies on the PES cancels the HLC terms. Finally, several small iodine-containing species (HOI, INO and ICN) and ClCN were characterized with standard G2 theory,¹¹ which employs ZPE calculated at the HF level and requires additional MP4/6-311G(d,p), MP4/6-311+G(d,p) and MP4/6-311G(2df,p) calculations (Table 4).

5. Discussion

5.1 Kinetics

As shown on Fig. 2, the present measurements for k_1 lie midway between the two earlier room-temperature determinations. Neither is obviously flawed, although the authors of the discharge - fast flow study³ (smaller k_1 value) suggested that the laser photolysis - resonance fluorescence study² (larger k_1 value) may have been affected by unrecognized secondary chemistry that consumed OH, even though the latter authors sought such effects by variation of the laser intensity but did not observe systematic changes in k_1 . The present data yield the first Arrhenius parameters for reaction 1, and the A factor found here, around 10^{-11} cm³ molecule⁻¹ s⁻¹, implies a fairly loose TS and is typical of A factors found for H abstraction reactions by OH from, for example, H₂ or CH₄.²³

The rate constant at the mean tropospheric temperature of 277 K, estimated from eq. 4 to be

$4.0 \times 10^{-14} \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$, is about 5 times that for $\text{OH} + \text{CH}_3\text{CCl}_3$, and thus the tropospheric lifetime of CF_3I with respect to OH attack is about 0.2 of that of CH_3CCl_3 , i.e. about 1.5 years.²⁴ This result may be compared to a photolytic lifetime of less than two days²⁵ which indicates that the major loss pathway for CF_3I in the troposphere is via photolysis.

Equation 4 could be extrapolated to combustion temperatures, but frequently it is found that rate expressions of the form $A T^n \exp(-B/T)$ are more realistic than simple Arrhenius expressions over wide temperature ranges. We have estimated $n = 1.5$ for reaction 1 from the vibrational frequencies of the TS and reactants, as outlined by Cohen,²⁶ and a fit of A and B to our data yields the following rate expression which is proposed for 280-2000 K:

$$k_1 = 2.9 \times 10^{-16} (T/\text{K})^{1.5} \exp(-960 \text{ K}/T) \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1} \quad (6)$$

This expression implies $k_1 \approx 10^{-11} \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$ at 1500 K so that attack by OH in a flame is rapid, and reaction 1 might therefore be a significant destruction pathway for CF_3I under combustion conditions.

5.2 Product channels

The products shown for reaction 1, $\text{CF}_3 + \text{HOI}$, are those that would be formed by I-atom abstraction by OH. The use of reaction 1 by Monks et al.⁶ as a source of HOI demonstrates that HOI production is a significant pathway, but other thermochemically plausible product channels might also be important. An alternative pathway is substitution



which is exothermic by 252 kJ mol^{-1} at 298 K.^{27,28} We were not able to find a TS for this reaction with attack by OH at the opposite side of CF_3I from the C-I bond, but did locate a TS for attack at the same face of CF_3I . Thus if substitution occurs it appears to preserve the CF_3 group without inversion. The HF/3-21G(d) TS structure has greatly extended C-I and C-O bonds, with lengths of 3.20 and $2.84 \times 10^{-10} \text{ m}$, respectively. The MP4/6-31G(d)//HF/3-21G(d) energy is 202 kJ mol^{-1} above $\text{CF}_3\text{I} + \text{OH}$, so that channel 1b has too high a barrier to contribute to the observed kinetics. As a check on the reliability of this barrier, the MP4/6-31G(d)//HF/3-21G(d) energy of the TS for reaction 1 was computed, and found to be 56 kJ mol^{-1} above $\text{CF}_3\text{I} + \text{OH}$. As discussed in the next section, this is about 70 kJ mol^{-1} too positive by comparison with more computationally intensive G2(MP2) results.

If a similar correction were applied to reaction 1b its barrier would still be very high, around 130 kJ mol⁻¹.

Ruscic and Berkowitz²⁹ noted that a possible product channel is



which is exothermic by 59 kJ mol⁻¹ at 298 K.^{27,28} We were not able to find a TS for this reaction, although this does not prove that channel 1c does not occur. However, the maximum MP4/6-31G(d) energy along a linear synchronous path connecting reactants and products is 574 kJ mol⁻¹ above the reactants. This is an upper limit to the barrier to this reaction if a TS exists, but suggests that channel 1c will be extremely slow.

Wine and McKee and coworkers have investigated adduct formation between halogenated methanes and halogen atoms and hydroxyl radicals. They observed chlorine atom adduct formation at low temperatures (around 250 K) and characterized adducts computationally using density functional theory.³⁰ We have analyzed



at the B3LYP/6-31G(d) level of theory and located an adduct with an I-O distance of 2.57×10^{-10} m, bound by 15 kJ mol⁻¹ at 0 K (including ZPE). MP2=full/6-31G(d) calculations did not reveal an adduct, and the G2(MP2) energy at the B3LYP/6-31G(d) geometry (Table 3) lies 1.5 kJ mol⁻¹ above CF₃I + OH. Therefore any adduct is not stable enough to be a significant sink for OH under our experimental conditions.

In summary, channels 1b-1d appear to be negligible under our conditions and the dominant products of reaction 1 are CF₃ + HOI, as assumed in the previous experimental kinetic studies.^{2,3}

5.3 Experimental thermochemistry and *ab initio* PES

The observation that the reaction OH + I₂ → I + HOI proceeds at the collision rate at 298 K implies this process cannot be endothermic,³¹ and therefore the bond dissociation enthalpy D₂₉₈(HO-I) must be at least D₂₉₈(I-I) = 151 kJ mol⁻¹, which combined with the known heats of formation²⁷ of I and OH implies Δ_fH₂₉₈(HOI) < -5 kJ mol⁻¹. Jenkin et al. analyzed earlier thermodynamic data^{32,33} to obtain -88 and -82 kJ mol⁻¹ for this quantity.³¹ Based on the idea that both reaction 1 and the process HONO + HOI → H₂O + INO₂ are rapid, Jenkin et al.³¹ deduced that -77 kJ mol⁻¹ ≥ Δ_fH₂₉₈(HOI) ≥ -

102 kJ mol⁻¹. The upper bound rests on the CF₃-I bond strength which Asher and Ruscic recently determined to high precision, $D_{298}(\text{CF}_3\text{-I}) = 227.2 \pm 1.3 \text{ kJ mol}^{-1}$.²⁸ This is more positive than the value of 223 kJ mol⁻¹ available to Jenkin et al.,³¹ and therefore raises the upper bound to $\Delta_f H_{298}(\text{HOI})$.

We can use our measured activation energy for reaction 1 to improve the estimates of $\Delta_f H_{298}(\text{HOI})$, based on the concept that the enthalpy change for a reaction is equal to the difference between the forward and reverse activation energies:

$$\Delta H = E_{a,1} - E_{a,-1} \quad (7)$$

Using the CF₃-I bond strength at 350 K (the bond strength varies by less than 0.1 kJ mol⁻¹ between 298 and 350 K),²⁷ a temperature that corresponds to the central 1/T value of the experiments, it is tempting to assume that the activation energy for the reverse reaction, -1, is given by $E_{a,-1} \geq 0$ and thus $D_{350}(\text{HO-I})$ must be at least 214.5 kJ mol⁻¹. However the reverse reaction could conceivably have a negative activation energy, perhaps as negative as -5 kJ mol⁻¹, and so we more conservatively derive $D_{350}(\text{HO-I}) \geq 209.5 \text{ kJ mol}^{-1}$ from experimental data alone. This result contradicts an empirical estimate by Zhang et al.³⁴ that $D_0(\text{HO-I}) = 189 \pm 2.5 \text{ kJ mol}^{-1}$ with an associated $\Delta_f H_0(\text{HOI}) = -42.7 \pm 2.5 \text{ kJ mol}^{-1}$ (which implies $\Delta_f H_{298}(\text{HOI}) = -46 \text{ kJ mol}^{-1}$). Our result also contradicts $\Delta_f H_0(\text{HOI}) \approx -36 \text{ kJ mol}^{-1}$ estimated empirically by Ruscic and Berkowitz.²⁹

Next we consider the PES for reaction -1 in more detail, using the G2(MP2) ab initio results (Table 3). These yield the enthalpy of the TS relative to CF₃ + HOI at 0 K, $E_{0,-1}^\ddagger = -13 \text{ kJ mol}^{-1}$. This negative value is consistent with, but does not prove, that there is no barrier to the reverse of reaction 1. This idea is tested in more detail on Fig. 4, where energies obtained at an approximate QCISD/6-311+G(3df,2p) level³⁵ are plotted as a function of the C-I distance along the reaction coordinate.³⁶ It may be seen there is no barrier to reaction -1, a result also derived from the low-level HF/6-31G(d) calculations discussed earlier but now confirmed at a much higher level of theory.

This conclusion implies an approximately zero activation energy for -1, and we take $E_{a,-1} = 0 \pm 5 \text{ kJ mol}^{-1}$. Since the enthalpies of formation of OH and I are well known,²⁷ eq. 8 is employed to estimate $\Delta_f H_{350}(\text{HOI}) = -69 \pm 5 \text{ kJ mol}^{-1}$, where the main source of uncertainty is $E_{a,-1}$ and the uncertainty in $E_{a,1}$ is negligible by comparison. The ab initio moments of inertia (Fig. 3) and frequencies (Table 2) are used to calculate the enthalpy function, $H_T - H_0$, which in combination with values for the elements in their reference states²⁷ yields $\Delta_f H(\text{HOI})$ at 298 and 0 K of -68 ± 5 and -63

$\pm 5 \text{ kJ mol}^{-1}$, respectively. We have calculated S , C_p and $H_T - H_{298}$ from 50 to 5000 K and fit the results with standard NASA polynomials³⁷ to obtain thermochemical data in a format suitable for combustion and atmospheric models (see Table 5). The corresponding O-I bond dissociation enthalpy is $D_{298}(\text{HO-I}) = 214 \pm 5 \text{ kJ mol}^{-1}$. The measured $\Delta_f H_{298}(\text{IO}) = 115.9 \pm 5.0 \text{ kJ mol}^{-1}$ from Bedjanian et al.³⁸ yields $D_{298}(\text{H-OI}) = 402 \pm 7 \text{ kJ mol}^{-1}$, comparable to the C-H bond strength in ethane.³⁹ Thus H-abstraction by IO from hydrocarbons is thermodynamically plausible and may need to be taken into account in models of iodine-mediated flame suppression.

5.4 *Ab initio thermochemistry*

The proposed thermochemistry of HOI is in reasonable accord with the earlier experimental limit values outlined above, but there are no measurements with which to compare the new result. The recent implementation of G2 theory for iodine compounds by Glukhovtsev et al.²⁰ enables an independent check of $\Delta_f H_0(\text{HOI})$. The G2 energy of HOI (Table 4), together with values for the atoms,^{11,20} yields ΔH_0 for atomization. However, this process involves considerable changes in the electronic structures of the atoms which might not be well accounted for, so we have also considered 7 other processes, listed in Table 6, which relate the unknown $\Delta_f H_0(\text{HOI})$ to known values for other small molecules. The computed ΔH_0 values derived from the G2 energies^{14,40} (see Table 4) for these reactions together with experimental values for $\Delta_f H_0$ of H, O, I, H_2O , HI, H_2S , ClCN, ICN, HOCl, HCl,²⁷ HNO ,⁴¹ INO ,⁴² CH_3I ⁴³ and HSI ⁴⁴ lead to the $\Delta_f H_0(\text{HOI})$ values listed in Table 6. The mean of the 8 values is -60 kJ mol^{-1} (standard deviation 6 kJ mol^{-1}) which is in good accord with our new experimental value of $-63 \pm 5 \text{ kJ mol}^{-1}$.

This computed $\Delta_f H_0(\text{HOI})$ is 15 kJ mol^{-1} more negative than the value recently obtained by Glukhovtsev et al.⁴⁵ also using G2 theory. The difference appears to arise about equally from their use of effective core potentials, by contrast to the all-electron basis sets employed here, and from their consideration of only the first working reaction of Table 5. This atomization step gives the most positive $\Delta_f H_0(\text{HOI})$ of all the processes analyzed, but we consider it the least reliable. Very recently Hassanzadeh and Irikura⁴⁶ used different working reactions and CCSD(T) theory to obtain $\Delta_f H(\text{HOI})$ at 0 and 298 K of -57.0 and $-61.7 \pm 7.0 \text{ kJ mol}^{-1}$, respectively, in excellent accord with the computed and experimentally-based values presented here.

6. Conclusions

The temperature dependence of the rate constant for the reaction $\text{OH} + \text{CF}_3\text{I}$ has been measured, and ab initio analysis suggests the dominant products are CF_3 and HOI . The measured Arrhenius parameters, together with an estimated activation energy $E_{a,1}$ for the reverse reaction supported by an ab initio analysis of the potential energy surface, yield the heat of formation of HOI . The main uncertainty in the present work lies in $E_{a,1}$. This experimental value for $\Delta_f H(\text{HOI})$ is confirmed by a direct computational estimate, and is more negative than recent empirical values. The temperature dependence of the rate constant suggests that OH attack is a minor contributor to the loss of CF_3I in the troposphere, but a potentially significant channel for CF_3I destruction under combustion conditions.

Acknowledgments

I thank Drs. Rajiv Berry (WPAFB), Jessie Yuan and Ashutosh Misra (UNT) for their invaluable assistance with this project.

References

1. Tapscott, R. E.; Skaggs, S. R.; Dierdorf, D. in *Halon Replacements* (Miziolek, A. W.; Tsang, W. Eds.); ACS Symp. Ser. 611: Washington, D.C., 1995.
2. Garraway, J.; Donovan, R. *J. Chem. Soc. Chem. Commun.* **1979**, 1108.
3. Brown, A. C.; Canosa-Mas, C. E.; Wayne, R. P. *Atmos. Environ.* **1990**, 24A, 361.
4. Wang, J. J.; Smith, D. J.; Grice, R. *J. Phys. Chem.* **1996**, 100, 6620.
5. Klaassen, J. J.; Lindner, J.; Leone, S. R. *J. Chem. Phys.* **1996**, 104, 7403.
6. Monks, P. S.; Stief, L. J.; Tardy, D. C.; Liebman, J. F.; Zhang, Z.; Kuo, S.-C.; Klemm, R. B. *J. Phys. Chem.* **1995**, 99, 16566.
7. Loomis, R. A.; Klaassen, J. J.; Lindner, J.; Christopher, P. G.; Leone, S. R. *J. Chem. Phys.* **1997**, 106, 3934.
8. Frey, J. G. in *Photodissociation Dynamics* (Balint-Kurti, G. G.; Law, M. M. Eds.); Collaborative Computational Project on Heavy Particle Dynamics: Daresbury Laboratory, Warrington, UK, 1994; p. 22.
9. Babushok, V.; Noto, T.; Burgess, D. R. F.; Hamins, A.; Tsang, W. *Combust. Flame* **1996**, 107, 351.
10. Solomon, S.; Garcia, R. R.; Ravishankara, A. R. *J. Geophys. Res.* **1994**, 99, 20,491.
11. Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, 94, 7221.
12. Shi, Y.; Marshall, P. *J. Phys. Chem.* **1991**, 95, 1654.
13. Ding, L.; Marshall, P. *J. Phys. Chem.* **1992**, 96, 2197.
14. Yuan, W.-J.; Misra, A.; Hommel, E.; Goumri, A.; Marshall, P. *to be published*.
15. (a) Marshall, P. *Computers Chem.* **1987**, 11, 219. (b) Marshall, P. *Computers Chem.* **1989**, 13, 103.
16. Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.

17. Foresman, J. B.; Frisch, A. E. *Exploring Chemistry with Electronic Structure Methods*; Gaussian: Pittsburgh, 1993.
18. Frisch, M. J.; Trucks, G. W.; Head-Gordon, M.; Gill, P. M. W.; Wong, M. W.; Foresman, J. B.; Johnson, B. G.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; DeFrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. A. *GAUSSIAN 92*; Gaussian: Pittsburgh, PA, 1992.
19. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *GAUSSIAN 94*; Gaussian: Pittsburgh, PA, 1995.
20. Glukhovtsev, M. N.; Pross, A.; McGrath, M. P.; Radom, L. *J. Phys. Chem.* **1995**, *103*, 1878.
21. Curtiss, L. A.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1995**, *103*, 4192.
22. Curtiss, L. A.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1993**, *98*, 1293.
23. Mallard, W. G.; Westley, F.; Herron, J. T.; Hampson, R. F. *NIST Chemical Kinetics Database - Ver. 6.0*; NIST Standard Reference Data: Gaithersburg, MD, 1994.
24. Chang, D. T.; Gurney, K. R.; Ko, M. K. W.; Kolb, C. E.; Nelson, D. D., Jr.; Rodriguez, J. M.; Weisenstein, D. K. in *Chemicals in the Atmosphere* (Chang, D. T.; Sze, N. D., Eds.); Atmospheric and Environmental Research: Cambridge, MA, 1992.
25. Solomon, S.; Burkholder, J. B.; Ravishankara, A. R.; Garcia, R. R. *J. Geophys. Res.* **1994**, *99*, 20929.
26. Cohen, N. *Int. J. Chem. Kinet.* **1989**, *21*, 909.
27. Chase Jr., M. W.; Davies, C. A.; Downey, Jr., J. R.; Frurip, D. J.; McDonald, R. A.; Syverud, A. N. *JANAF Thermochemical Tables*, 3rd Ed., *J. Phys. Chem. Ref. Data* **1985**, *14*, Suppl. 1.
28. Asher, R. L.; Appelman, E. H.; Tilson, J. L.; Litorja, M.; Berkowitz, J.; Ruscic, B. *J. Chem. Phys.* **1997**, *106*, 9111.

29. Ruscic, B.; Berkowitz, J. *J. Chem. Phys.* **1994**, *101*, 7795.
30. Piety, C. A.; Nicovich, J. M.; Ayhems, Y. V.; Estupinan, E.; Soller, R.; McKee, M. L.; Wine, P. H. *14th International Symposium on Gas Kinetics*, paper D6, 1996.
31. Jenkin, M. E.; Clemitshaw, K. C.; Cox, R. A. *J. Chem. Soc. Faraday Trans. 2* **1984**, *80*, 1633.
32. Gelles, E. *Trans. Faraday Soc.* **1951**, *47*, 1158.
33. Garisto, F. *Thermochim. Acta* **1983**, *63*, 251.
34. Zhang, Z.; Monks, P. S.; Stief, L. J.; Liebman, J. F.; Huie, R. E.; Kuo, S.-C.; Klemm, R. B. *J. Phys. Chem.* **1996**, *100*, 63.
35. Derived as in eq. 5 except that the triples contribution to the QCI energy and the ZPE were neglected.
36. The reaction coordinate was obtained by fixing the C-I distance and reoptimizing the structure. When the C-I distance exceeded 2.7×10^{-10} m the geometry had relaxed essentially to that of $\text{CF}_3 + \text{HOI}$ and so for subsequent points only the C-I distance was varied, with the rest of the structure held constant.
37. Burcat, A. in *Combustion Chemistry* (Gardiner, W. C., Jr.; Ed.) Springer-Verlag: New York, NY, 1984; Ch. 8 and Appendix A.
38. Bedjanian, Y.; Le Bras, G.; Poulet, G. *J. Phys. Chem. A* **1997**, *101*, 4088.
39. Berkowitz, J.; Ellison, G. B.; Gutman, D. *J. Phys. Chem.* **1994**, *98*, 2744.
40. Hommel, E.; Misra, A.; Marshall, P. *to be published*.
41. Anderson, W. R. *personal communication*, **1996**.
42. Hippler, H.; Luther, K.; Teitelbaum, H.; Troe, J. *Int. J. Chem. Kinet.* **1977**, *9*, 917.
43. Lias, S. G.; Bartmess, J. E.; Liebman, J. F.; Holmes, J. L.; Levin, R. D.; Mallard, W. G. *Gas-Phase Ion and Neutral Thermochemistry, J. Phys. Chem. Ref. Data* **1988**, *17*, Suppl. 1.
44. Hwang, R. J.; Benson, S. W. *J. Amer. Chem. Soc.* **1979**, *101*, 2615.
45. Glukhovtsev, M. N.; Pross, A.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 3498.
46. Hassanzadeh, P.; Irikura, K. K. *J. Phys. Chem. A* **1997**, *101*, 1580.

Table 1. Summary of experimental measurements for the reaction $\text{CF}_3\text{I} + \text{OH}$.

T, K	P, mbar	τ_{res} , s	F, J	$[\text{H}_2\text{O}]$, 10^{15} cm^{-3}	$[\text{CF}_3\text{I}]_{\text{max}}$, 10^{14} cm^{-3}	$k_{1,\text{eff}} \pm \sigma_{k1,\text{eff}}$, $10^{-13} \text{ molecule}^{-1} \text{ cm}^3 \text{ s}^{-1}$
281	68.2	1.6	5.00	3.8	12.1	1.96 ± 0.14
281	54.1	1.0	5.00	2.3	8.8	1.74 ± 0.04
281	89.8	1.7	5.00	3.1	10.9	1.67 ± 0.06
281	68.2	1.6	4.05	3.8	12.1	1.71 ± 0.10
281	54.1	1.0	4.05	2.3	8.8	1.52 ± 0.05
281	89.8	1.7	4.05	3.1	10.9	1.29 ± 0.08
281	68.2	1.6	2.50	3.8	10.0	1.25 ± 0.09
281	54.1	1.0	2.50	2.3	8.8	1.01 ± 0.06
281	89.8	1.7	2.50	3.1	10.9	1.07 ± 0.06
281						0.43 ± 0.01^a
292	75.2	1.7	5.00	2.8	17.8	1.39 ± 0.10
292	75.3	1.7	5.00	2.2	13.9	2.18 ± 0.15
292	71.0	1.6	5.00	4.0	12.7	2.10 ± 0.09
292	63.9	1.4	5.00	3.3	15.6	1.59 ± 0.13
292	68.0	1.5	4.05	2.6	22.1	1.48 ± 0.06
292	59.1	1.0	4.05	1.7	16.9	1.44 ± 0.05
292	71.1	1.6	4.05	2.8	14.3	1.88 ± 0.13
292	72.1	1.6	4.05	4.7	13.5	1.90 ± 0.12
292	75.2	1.7	2.50	2.8	17.9	1.26 ± 0.07

292	69.6	1.6	2.50	2.6	16.3	1.16 ± 0.08
292	64.0	1.4	2.50	3.3	15.2	1.20 ± 0.04
292	62.7	1.5	2.50	2.3	14.9	1.17 ± 0.02
292						0.58 ± 0.08^a
338	83.2	1.6	5.00	2.6	12.3	2.23 ± 0.13
335	67.5	1.1	5.00	2.0	8.5	2.21 ± 0.11
338	115.7	2.3	4.05	4.3	16.2	2.18 ± 0.12
338	76.5	1.1	4.05	2.2	9.5	1.77 ± 0.09
338	83.2	1.6	4.05	2.6	12.3	2.11 ± 0.15
338	83.2	1.6	2.50	2.6	12.3	1.84 ± 0.13
335	67.5	1.1	2.50	2.0	8.5	1.49 ± 0.08
337						0.99 ± 0.10^a
391	100.6	1.7	5.00	3.4	11.4	3.38 ± 0.19
390	68.5	1.0	5.00	2.1	6.9	3.03 ± 0.14
391	100.6	1.7	4.05	3.4	11.4	2.55 ± 0.30
390	68.5	1.0	4.05	2.1	6.9	2.69 ± 0.07
391	100.6	1.7	2.50	3.4	11.4	2.52 ± 0.18
390	68.5	1.0	2.50	2.1	6.9	2.26 ± 0.17
391						1.68 ± 0.39^a
445	92.0	1.3	5.00	3.0	2.0	7.06 ± 0.27
443	89.0	1.8	5.00	3.4	8.1	4.95 ± 0.18
442	59.5	1.2	5.00	2.4	7.0	4.41 ± 0.13
445	92.0	1.3	4.05	3.0	2.0	6.45 ± 0.55

443	89.0	1.8	4.05	3.4	8.1	4.34 ± 0.17
442	59.5	1.2	4.05	2.4	7.0	3.94 ± 0.18
445	92.0	1.3	2.50	3.0	2.0	4.48 ± 0.17
443	89.0	1.8	2.50	3.4	6.0	3.92 ± 0.23
442	59.5	1.2	2.50	2.4	7.0	3.64 ± 0.19
443						3.38 ± 0.41^a

^alinear extrapolation of k_{eff} to zero flash energy for each temperature and 1σ uncertainty of the intercept.

Table 2. Fundamental vibrational frequencies ν_0 of stationary points on the $\text{CF}_3\text{I} + \text{OH}$ PES (cm^{-1}), calculated at the HF and MP2 levels of theory with the 6-31G(d) basis set.^a

CF_3I			OH			CF_3			HOI			TS
HF	MP2	Expt. ^b	HF	MP2	Expt. ^b	HF	MP2	Expt. ^b	HF	MP2	Expt. ^c	MP2
263 (2)	265 (2)	260 (2)	3570	3607	3570	491 (2)	486 (2)	500 (2)	589	563	575	149 i
278	278	284				677	680	701	1088	1116	1068	24
521 (2)	514 (2)	539 (2)				1086	1086	1090	3659	3577	3626	94
726	718	742				1286 (2)	1271 (2)	1259 (2)				147
1070	1057	1074										171
1243 (2)	1219 (2)	1185 (2)										208
												490
												497
												626
												796
												959
												1091
												1242
												1261
												3608

^aHF values scaled by 0.8929 and MP2 values scaled by 0.9646. Doubly degenerate modes are indicated by (2).

^bRef. 27. ^cRef. 5.

Table 3. G2(MP2) energies for stationary points on the OH + CF₃I PES, in hartrees.^a

Species	QCISD(T)/ 6-311G(d,p)	MP2/6-311 +G(3df,2p)	MP2/ 6-311G(d,p)	HLC	ZPE	G2(MP2)
OH	-75.58921	-75.61742	-75.57276	-0.01519	0.00822	-75.64084
CF ₃ I	-7253.96099	-7254.18589	-7253.92435	-0.08000	0.01378	-7254.28875
CF ₃ I-OH [‡] TS	-7329.53285	-7329.79873	-7329.47649	-0.09519	0.02554	-7329.92474
HOI	-6992.56705	-6992.65970	-6992.54346	-0.03500	0.01197	-6992.70632
CF ₃	-336.96752	-337.14009	-336.94227	-0.06019	0.01203	-337.21350
CF ₃ IOH adduct	-7329.54763	-7329.80225	-7329.49104	-0.09519	0.02503 ^b	-7329.92900

^aBased on scaled MP2 frequencies and HLC = -0.00019 ν_a - 0.00481 ν_b . 1 hartree \approx 2625 kJ mol⁻¹.

^bBased on unscaled B3LYP/6-31G(d) frequencies.

Table 4. G2 energies of ancillary molecules used to estimate $\Delta_f H_0(\text{HOI})$.^a

Species	MP4/6-311G(d,p)	ZPE	$\Delta E(+)$	$\Delta E(2df)$	$\Delta E(\text{QCI})$	Δ	$E_0(\text{G2})$
HOI	-6992.56772	0.01215	-0.01045	-0.09163	0.00067	-0.02483	-6992.71681
INO	-7046.58764	0.00591	-0.00725	-0.11020	0.00943	-0.02678	-7046.76153
ICN	-7009.55409	0.00784	-0.00443	-0.09583	0.00477	-0.02382	-7009.70557
CICN	-552.27617	0.00873	-0.00671	-0.10834	0.00466	-0.00687	-552.42469

^aIn hartrees.

Table 5. Thermochemistry of HOI and NASA polynomials.

$\Delta_f H_0(\text{HOI}) = -63 \pm 5 \text{ kJ mol}^{-1}$	$\Delta_f H_{298}(\text{HOI}) = -68 \pm 5 \text{ kJ mol}^{-1}$	
$C_{p,298} = 38.5 \text{ J K}^{-1} \text{ mol}^{-1}$	$S_{298} = 255.4 \text{ J K}^{-1} \text{ mol}^{-1}$	$H_{298}-H_0 = 10.4 \text{ kJ mol}^{-1}$
$C_p/R = a_1 + a_2 T + a_3 T^2 + a_4 T^3 + a_5 T^4$		
$(\Delta_f H_{298} + H_T - H_{298})/RT = a_1 + a_2 T/2 + a_3 T^2/3 + a_4 T^3/4 + a_5 T^4/5 + a_6/T$		
$S/R = a_1 \ln T + a_2 T + a_3 T^2/2 + a_4 T^3/3 + a_5 T^4/4 + a_7$		
0.26354949E+01 0.52934503E-02-0.25414167E-05 0.53307325E-09-0.40591408E-13		
-0.89380615E+04 0.14730852E+02 0.39206498E+01 0.67459485E-04 0.12938654E-04		
-0.20159172E-07 0.91124192E-11-0.94277812E+04 0.79484367E+01 ^a		

^aFirst seven numbers are $a_1 - a_7$ for 1000-5000 K, followed by $a_1 - a_7$ for 50-1000 K.

Table 6. G2 thermochemistry of working reactions used to calculate $\Delta_f H_0(\text{HOI})$, kJ mol^{-1} .

Reaction	ΔH_0	$\Delta_f H_0(\text{HOI})^a$
$\text{HOI} \rightarrow \text{H} + \text{O} + \text{I}$	623.6	-53.6
$\text{HOI} + \text{H}_2 \rightarrow \text{H}_2\text{O} + \text{HI}$	-150.4	-60.0
$\text{HOI} + \text{HCl} \rightarrow \text{HOCl} + \text{HI}$	105.0	-55.8
$\text{HOI} + \text{HI} \rightarrow \text{H}_2\text{O} + \text{I}_2$	-140.4	-61.6
$\text{HOI} + \text{ClCN} \rightarrow \text{ICN} + \text{HOCl}$	71.8	-54.0
$\text{HOI} + \text{HNO} \rightarrow \text{INO} + \text{H}_2\text{O}$	-161.1	-62.7
$\text{HOI} + \text{H}_2\text{S} \rightarrow \text{HSI} + \text{H}_2\text{O}$	-103.0	-73.0
$\text{HOI} + \text{CH}_4 \rightarrow \text{CH}_3\text{I} + \text{H}_2\text{O}$	-90.2	-56.8

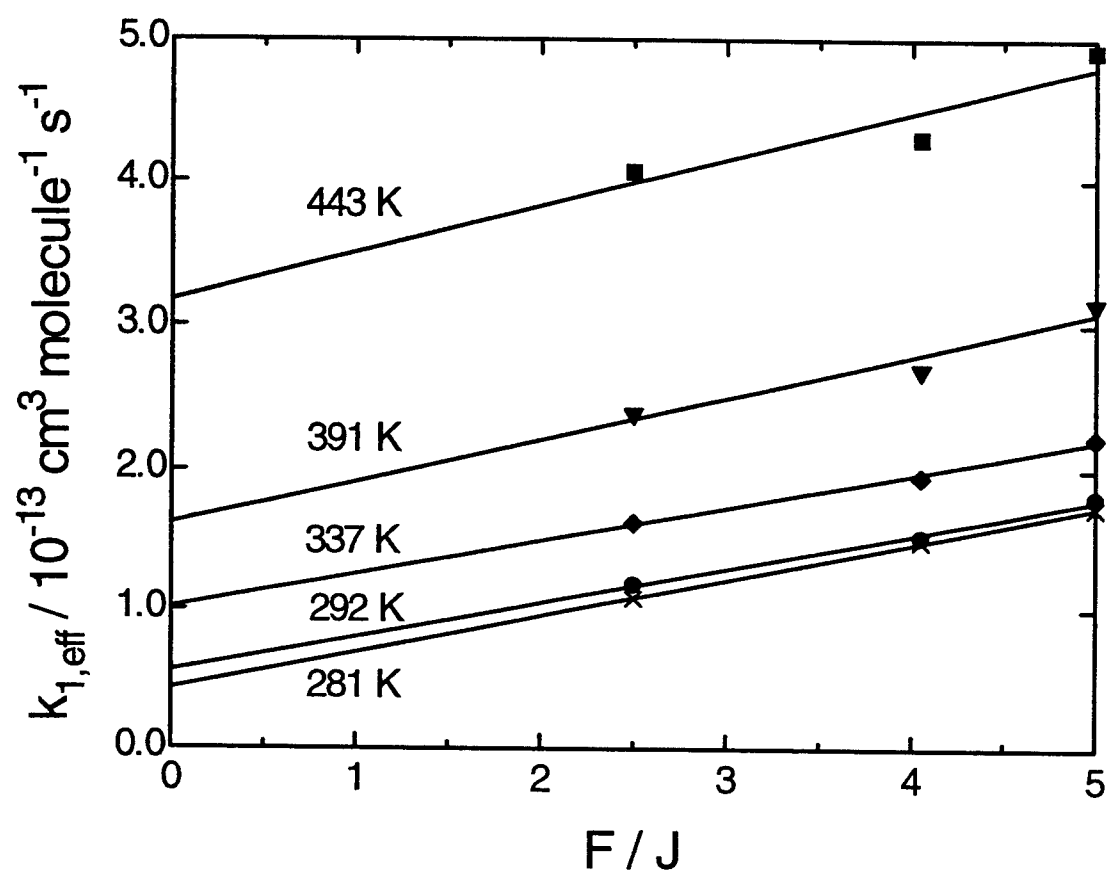
Figure captions

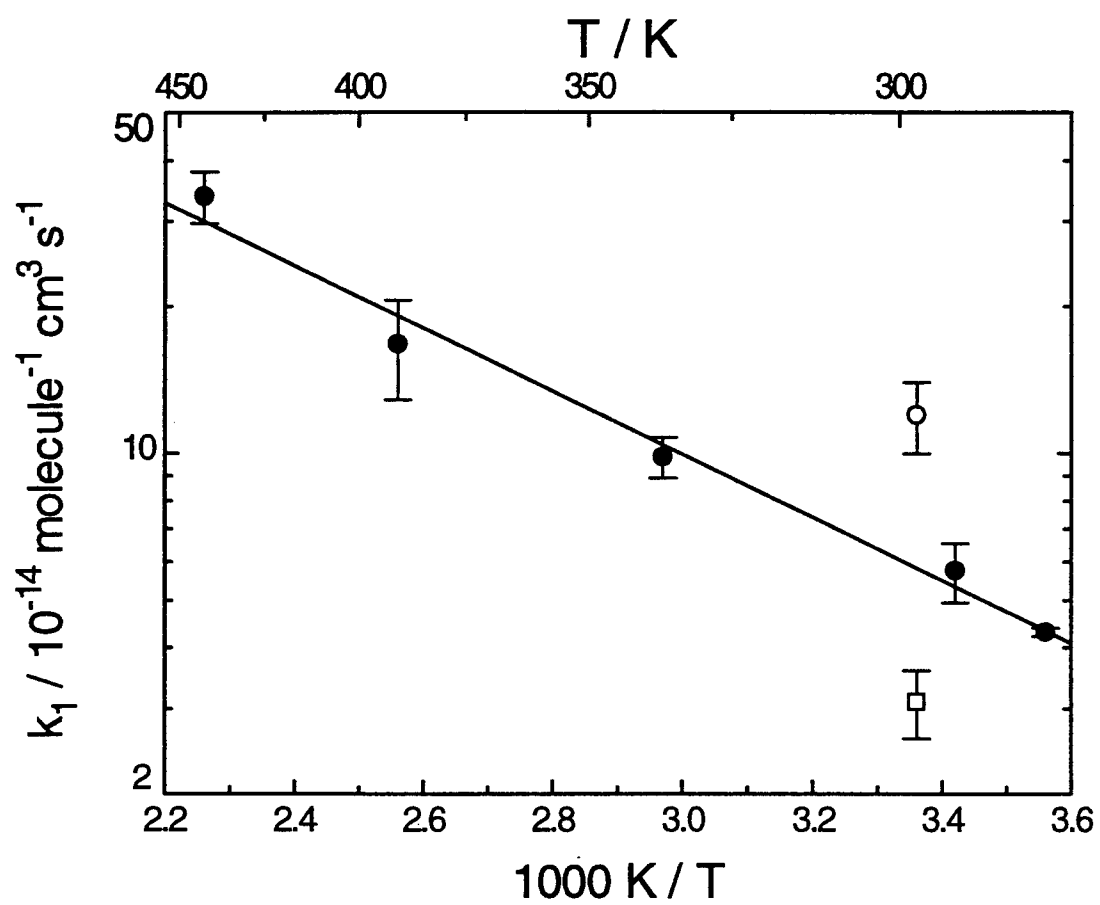
Fig. 1 Extrapolations of the effective rate constant for $\text{OH} + \text{CF}_3\text{I}$, k_{eff} , to zero radical concentration (flash lamp energy $F = 0$) at each temperature. The points represent weighted means of measurements at the same F .

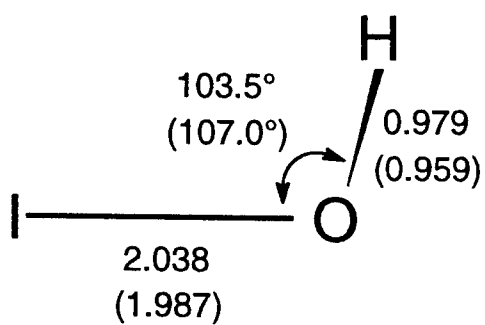
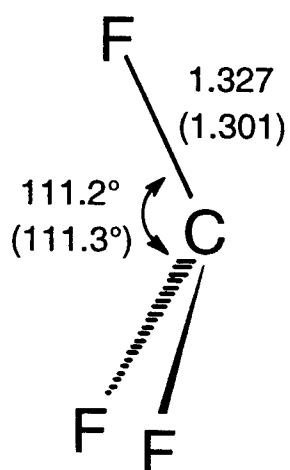
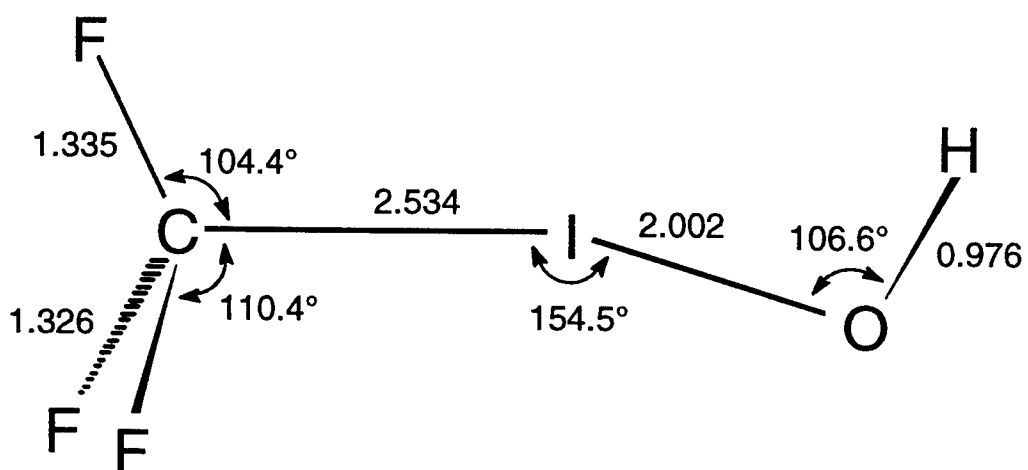
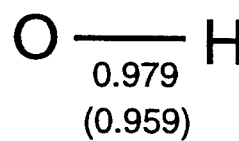
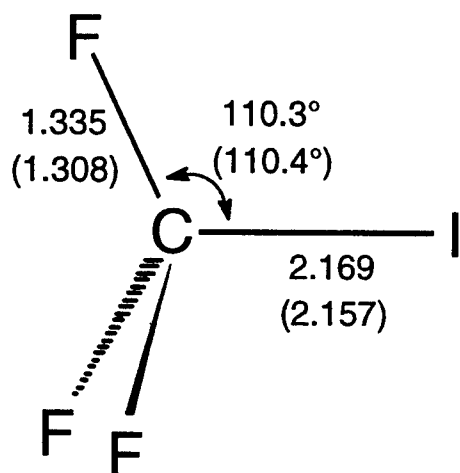
Fig. 2 Arrhenius plot for $\text{OH} + \text{CF}_3\text{I}$. (●) present work; (○) ref. 2; (□) ref. 3. Error bars represent $\pm 1\sigma$ precision.

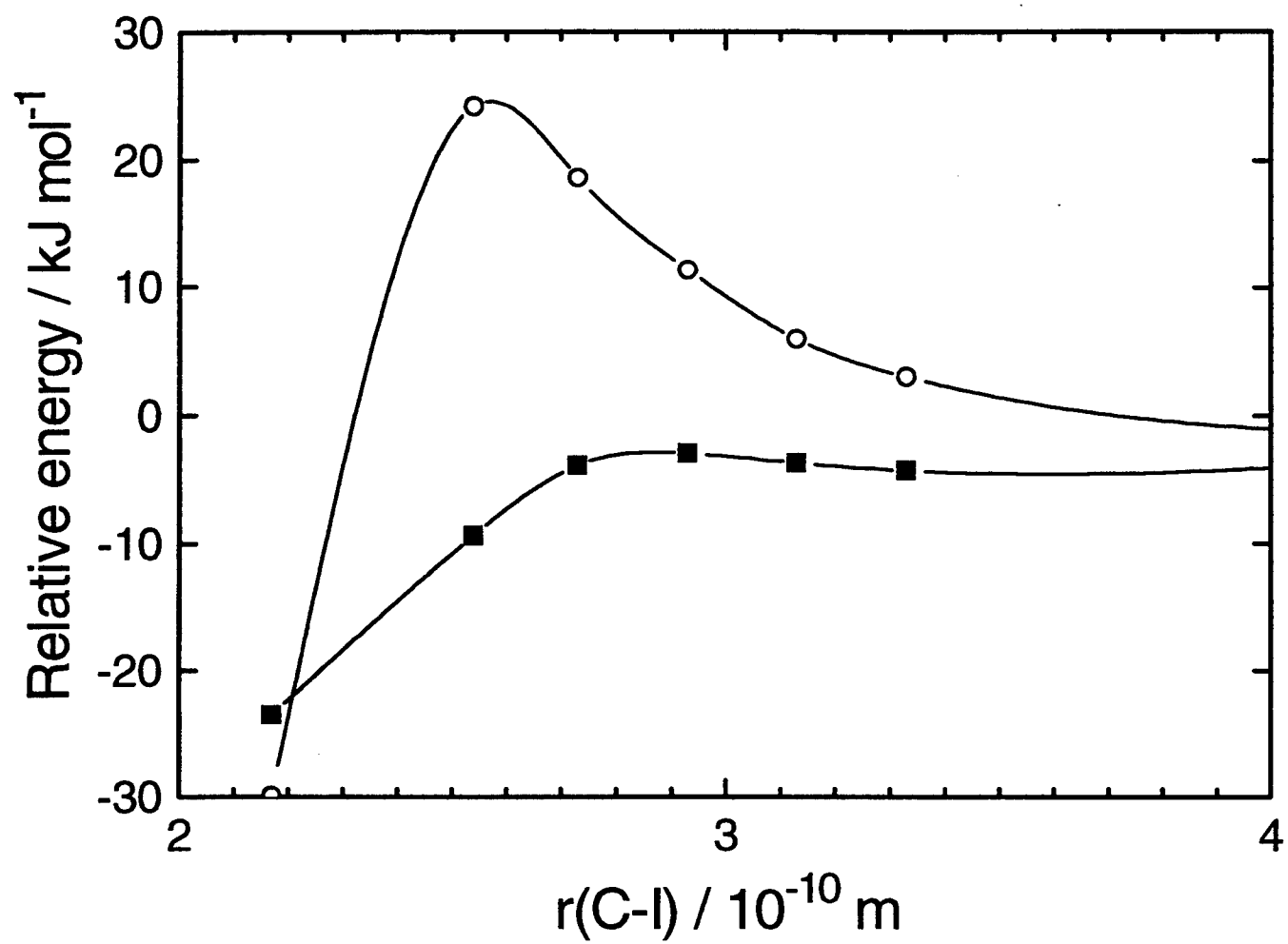
Fig. 3 Geometries of reactants, transition state and products for the $\text{OH} + \text{CF}_3\text{I}$ PES. Angles in degrees and bond lengths in 10^{-10} m. MP2=full/6-31G(d) theory employed, with HF/6-31G(d) values in parentheses. CF_3 and CF_3I have C_{3v} symmetry. Dihedral angles in the TS: CIOH , 81.9° ; FCIO (long C-F bond), 175.6° ; FCIO (short C-F bonds), 56.7° and -65.9° .

Fig. 4 Relative energies along the reaction coordinate for $\text{OH} + \text{CF}_3\text{I} \rightarrow \text{HOI} + \text{CF}_3$ computed at (○) MP2=full/6-31G(d) and (■) approximate (see footnote 36) QCISD/6-311+G(3df,2p) levels of theory. The energy zero is set at $\text{HOI} + \text{CF}_3$, where $r(\text{C-I}) = \infty$. $r(\text{C-I}) = 2.17 \times 10^{-10}$ m in the reactants and 2.53×10^{-10} m at the MP2=full/6-31G(d) TS.









Hui Meng
Report not available at time of publication.

**BAND-GAP CALCULATIONS ON SQUARATE-CONTAINING CONJUGATED OLIGOMERS
FOR THE PREDICTION OF CONDUCTIVE AND NON-LINEAR
OPTICAL PROPERTIES OF POLYMERIC MATERIALS**

Douglas J. Miller, PhD
Professor of Chemistry
Science & Mathematics Department

Cedarville College
PO Box 601
Cedarville, OH 45314-0601

Final Report for:
Summer Research Extension Program
Wright Laboratory/Materials Directorate

Sponsored by:
Air Force Office of Scientific Research
Bolling AFB, Washington DC

and

Wright Laboratory/Materials Directorate

March 1998

**BAND-GAP CALCULATIONS ON SQUARATE-CONTAINING CONJUGATED OLIGOMERS
FOR THE PREDICTION OF CONDUCTIVE AND NON-LINEAR
OPTICAL PROPERTIES OF POLYMERIC MATERIALS**

Douglas J. Miller, PhD
Professor of Chemistry
Science & Mathematics Department
Cedarville College

ABSTRACT

Attempts were made to calculate the band gaps of conjugated oligomers containing squarate moieties and five-membered rings which incorporated nitrogen atoms (pyrrole-type structure). Computational software (Gaussian, GAMESS, and HyperChem) was utilized in this work. Extrapolation of the results for the oligomers to the polymeric material was made in order to ascertain whether it might possess desired conductive and nonlinear optical properties. (Figures and tables are placed after the footnotes page.)

**BAND-GAP CALCULATIONS ON SQUARATE-CONTAINING CONJUGATED OLIGOMERS
FOR THE PREDICTION OF CONDUCTIVE AND NON-LINEAR
OPTICAL PROPERTIES OF POLYMERIC MATERIALS**

Douglas J. Miller, PhD

INTRODUCTION

Conjugated polymers are important for their applications in electronic technologies.¹ The band-gap is a key material property for (semi)conductive and nonlinear optical properties, and much of the current focus is on designing new polymers having small band gaps.² Typical undoped conjugated polymers have band gaps in the 1.5 - 3.5 eV range, so polymers possessing values less than or equal to 1.0 eV represent important candidate materials.

We recently proposed³ a scheme whereby low band gap materials could be designed. The underlying principle is that of symmetric ground states. Briefly, these are polymers in which two equivalent, isoenergetic, isostructural resonance forms contribute equally to the real structure. Efforts begun during the 1996 Summer Research Program extended this work to systems having an all-carbon backbone, a negative charge, and a very rigid structure. The current research extended this work to a system which contained both the squarate group and five membered rings which incorporated nitrogen atoms. The squarate group has been in molecules of recent interest for their nonlinear optical properties.⁴ All of the studied oligomers had rigid structures and an overall neutral charge.

METHODS AND DISCUSSION

Electronic properties for the polymer of interest were extrapolated from values obtained from computational software (Gaussian^{5,6}, GAMESS⁷, and HyperChem) for oligomers. The potentially conducting polymer was an ABAB copolymer in which A was the 1,3-squarate unit and B was the 2,5'-bipyrrole unit. The shortest oligomer, shown in Figure 1 and designated P2S, did not contain a bipyrrole; but all higher oligomers did. The oligomers are designated as PnS or PnB, where n refers to the total number of pyrrole rings, and the S indicates that a squarate moiety is in the center of the oligomers while a B indicates that a bipyrrole occupies the central portion of the molecule. The oligomer designated P4B, shown in Figure 2, is generated by creating a "dimer" of the P2S structure. P6S is generated by adding a third P2S unit.

All of the oligomers have C_{2h} symmetry and are terminated by single pyrrole rings. The two slightly structurally different series of oligomers thus generated (P2S, P6S, P10S, ... and P4B, P8B, P12B, ...) should, in the infinite limit, represent the same polymer. The oligomers have a pattern of alternating double bonds, possess rigid planar structures, and are, overall, electrically neutral. Each mer unit (pyrrole-squarate-pyrrole), however, has a formal -1 charge on an oxygen of the squarate and a formal +1 charge on the nitrogen atom of one of the pyrrole rings.

Quantum chemistry software was first used to find optimized geometries using semiempirical (AM1) parameters. Rough z-matrices were used as input for geometry optimizations in the Gaussian 94W program (running on a PC). The geometries were then fully optimized within C_{2h} symmetry at the AM1

theory level using the GAMESS program. The final geometries were then used as structural input for electronic calculations using GAMESS or spectroscopic calculations using HyperChem.

The band-gap for the polymer was determined by extrapolating, to an infinite length, the band-gap values obtained for the oligomers. The band-gap for each oligomer was calculated by finding the difference between the energy of the ground state (singlet) and the lowest lying excited state (triplet). These energies were determined by Hartree-Fock methods in GAMESS. The singlet states were calculated at the RHF/6-31G* level while the triplet states were calculated at the ROHF/6-31G* level. All calculations were performed at the ground state geometry obtained by the AM1 optimizations described above. We noted that the second order SCF (SOSCF) energies did not always converge correctly, so all results reported were obtained with SOSCF turned off.

Vertical electron affinities for the oligomers were determined as the difference in energy between the ground state (singlet) and the monoanion (doublet) obtained by adding an electron to the neutral oligomers. The monoanion (doublet) energy was calculated at the ROHF/6-31G* level of theory. This was done in order to estimate the stability of the neutral species in the real world.

The band-gap for the polymer was also obtained by finding the energy difference between the the ground state (singlet) and first excited state (also singlet in this case) for each oligomer and extrapolating the results to infinite chain length. These energies were obtained by the ZINDO/S method as implemented in HyperChem (Release 5.01 for Windows95). Each CI

calculation incorporated the 15 highest energy, occupied orbitals and the 15 lowest unoccupied orbitals (a 15x15 CI). The SCF Convergence Limit was set to 0.0001, Accelerate Convergence was set to "off", and the Pi-Pi overlap was changed from the default to 0.585 as recommended for organic materials.⁸ All calculations employed the AM1 geometries obtained above. These HyperChem calculations were carried out to compare to the results obtained from GAMESS and to give impressions of trends if GAMESS results were not obtainable.

RESULTS AND CONCLUSION

The energies, in Hartrees as determined in GAMESS, of the ground states (singlets) and the excited states (triplets) of various oligomers are listed in Table 1. Ground state energies were determined for all oligomers up through P12B. Determination of the excited states became problematic for P10S and P12B, and efforts are continuing.

Band-gap values for the oligomers, as calculated from both the GAMESS results and the HyperChem ZINDO/S computations, are listed in Table 2. Figure 3 illustrates the four band-gap data points (connected using a cubic B-spline curve) obtained from GAMESS. The x-axis units are $1/n$ where n is the number of pyrrole rings in the oligomer. The graph shows that the value of the band-gap decreases with increasing length (smaller $1/n$ values) of the oligomer. This graph suggests that the band-gap for the polymer ($1/n$ near zero) is below 0.60 eVs. More data points are required to more clearly establish the trend, especially since the PnS and PnB series will likely take somewhat different paths to the same ultimate band-gap value for the polymer. Connecting the band-gap values for P4B and P8B with a straight

line (see Figure 3) leads to a band-gap of 0.488 eVs at infinite length. A similar line drawn through the band-gap values for P2S and P6S leads to a band-gap value of 0.531 eVs at infinite length. Figure 5 illustrates the relationship of band-gap values, as determined by HyperChem, for all six oligomers (P2S, P4B, P6S, P8B, P10S, and P12B). Again the connected points appear to be directed toward the neighborhood of 0.60 eVs. A "best" straight line drawn through these data points, minus the one for P2S, leads to a band-gap of 0.552 eVs at infinite length (see Figure 6). The band-gap value for P2S was left out of Figure 6 since P2S is the most likely oligomer to exhibit the effects of limited length and the effects due to being end-capped by single pyrroles.

All of the results point toward a band-gap value of less than 0.60 eVs for the candidate polymer. This could indicate that it would be worthwhile to attempt to synthesize the material in order to study its conductive and non-linear optical properties. The electron affinities (Table 3) point to a potential instability, relative to reduction, of this polymer in the real world. Polymers with similar structures that have other substituents attached would also be worthy candidates for computational investigation.

DJM:ljm

REFERENCES

- (1) Gorman, C. B.; Grubbs, R. H. In *Conjugated Polymers*; Bredas, J. L.; Silbey, R., Eds.; Kluwer Academic Publishers: Boston, 1991; pp. 1-48.
- (2) Kertesz, M.; Ashertehrani A. *Macromolecules* 1996, 29(3), 940-945.
- (3) Unpublished Results.
- (4) Dirk, C. W.; Herndon, W. C.; Cervantes-Lee, F.; Seinau, H.; Martinez, S.; Kalamegham, P.; Tan, A.; Campos, G.; Velez, M.; Zyss, J.; Ledoux, I.; Cheng, L. J. *Am. Chem. Soc.* 1995, 117(8), 2214-2225.
- (5) Gaussian 92, Revision A, M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 1992.
- (6) Gaussian 94, Revision A, M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. A. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanayakkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 1995.
- (7) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* 1993, 14, 1347-1363.
- (8) Ridley, J. E.; Zerner, M. C. *Theoret. Chim. Acta* 1976, 42, 223-236.

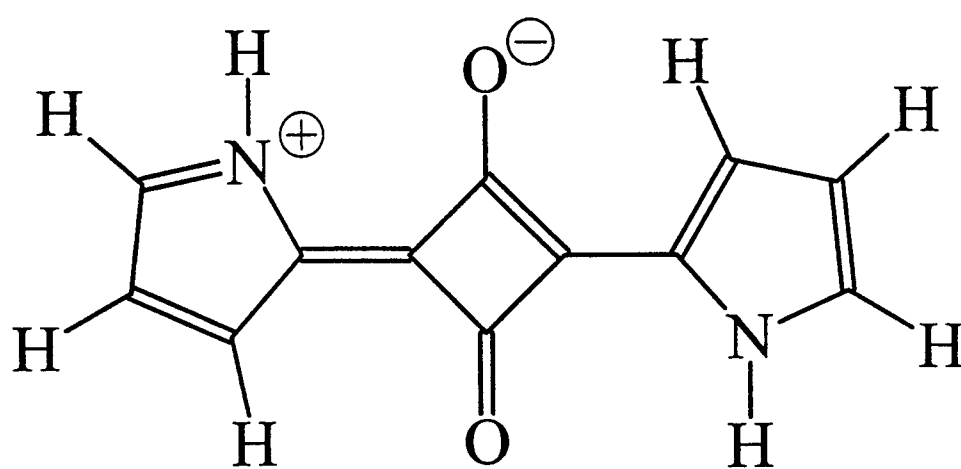


Figure 1
P2S Molecule

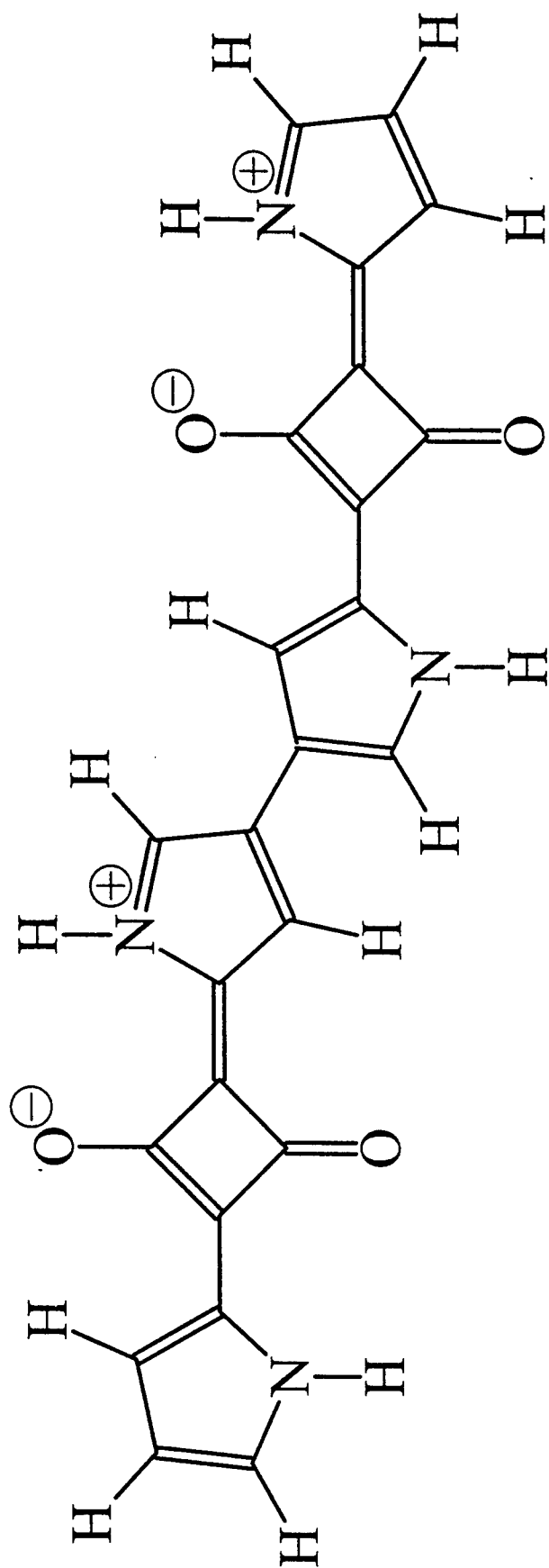


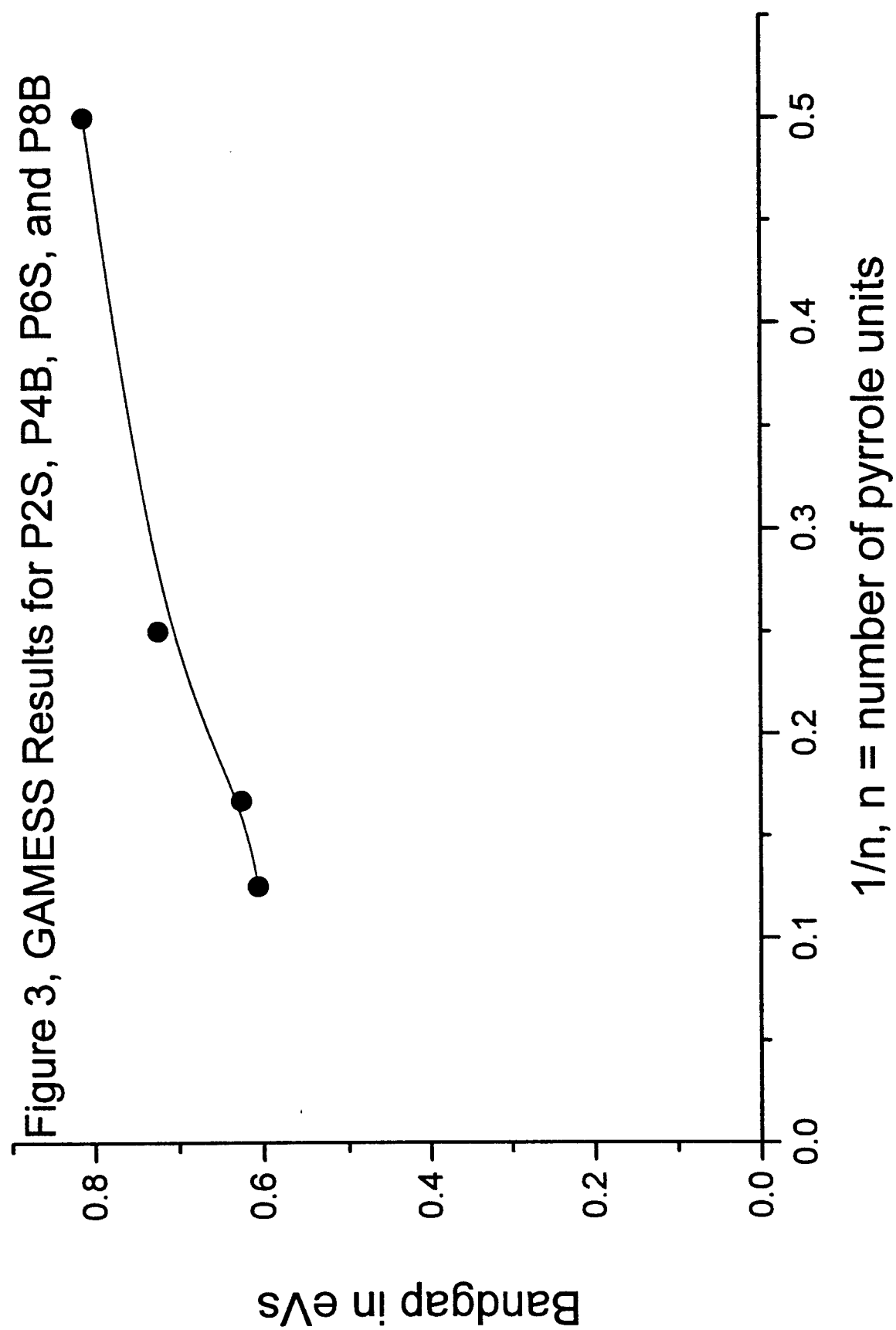
Figure 2
P4B Molecule

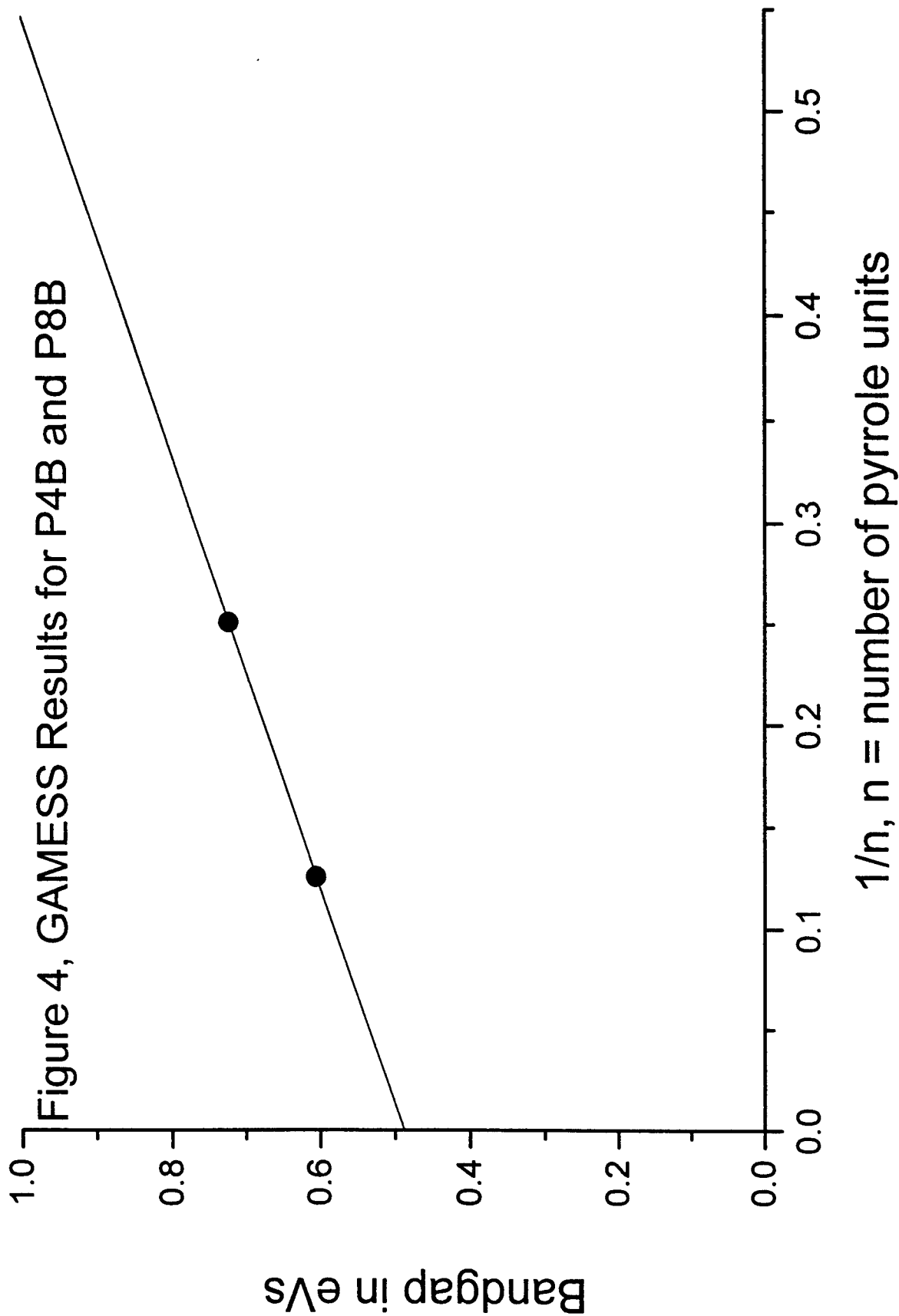
Table 1: Energies (in Hartrees) as Determined by GAMESS

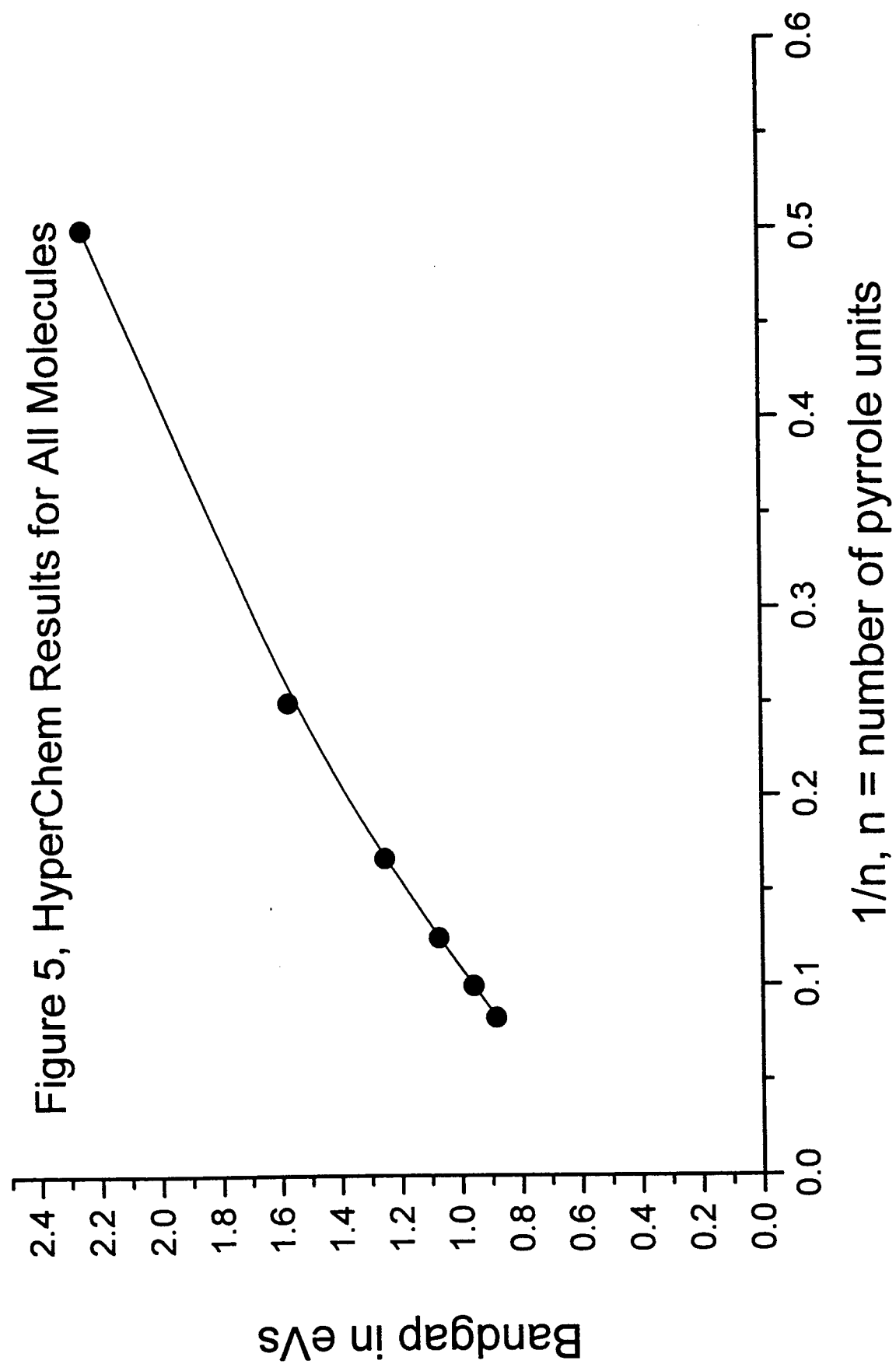
Oligomer Designation	Ground State (Singlet)	Excited State (Triplet)
P2S	-717.6080634653	-717.5782832373
P4B	-1434.0716942861	-1434.0451194595
P6S	-2150.5352758266	-2150.5123118523
P8B	-2866.9987431638	-2866.9764854880
P10S	-3583.4620287117	-
P12B	-4299.9245800857	-

Table 2: Band-gap Energies (in eVs)

Oligomer Designation	From GAMESS	From HyperChem
P2S	0.8103	2.241
P4B	0.7231	1.569
P6S	0.6248	1.250
P8B	0.6056	1.072
P10S	-	0.958
P12B	-	0.885







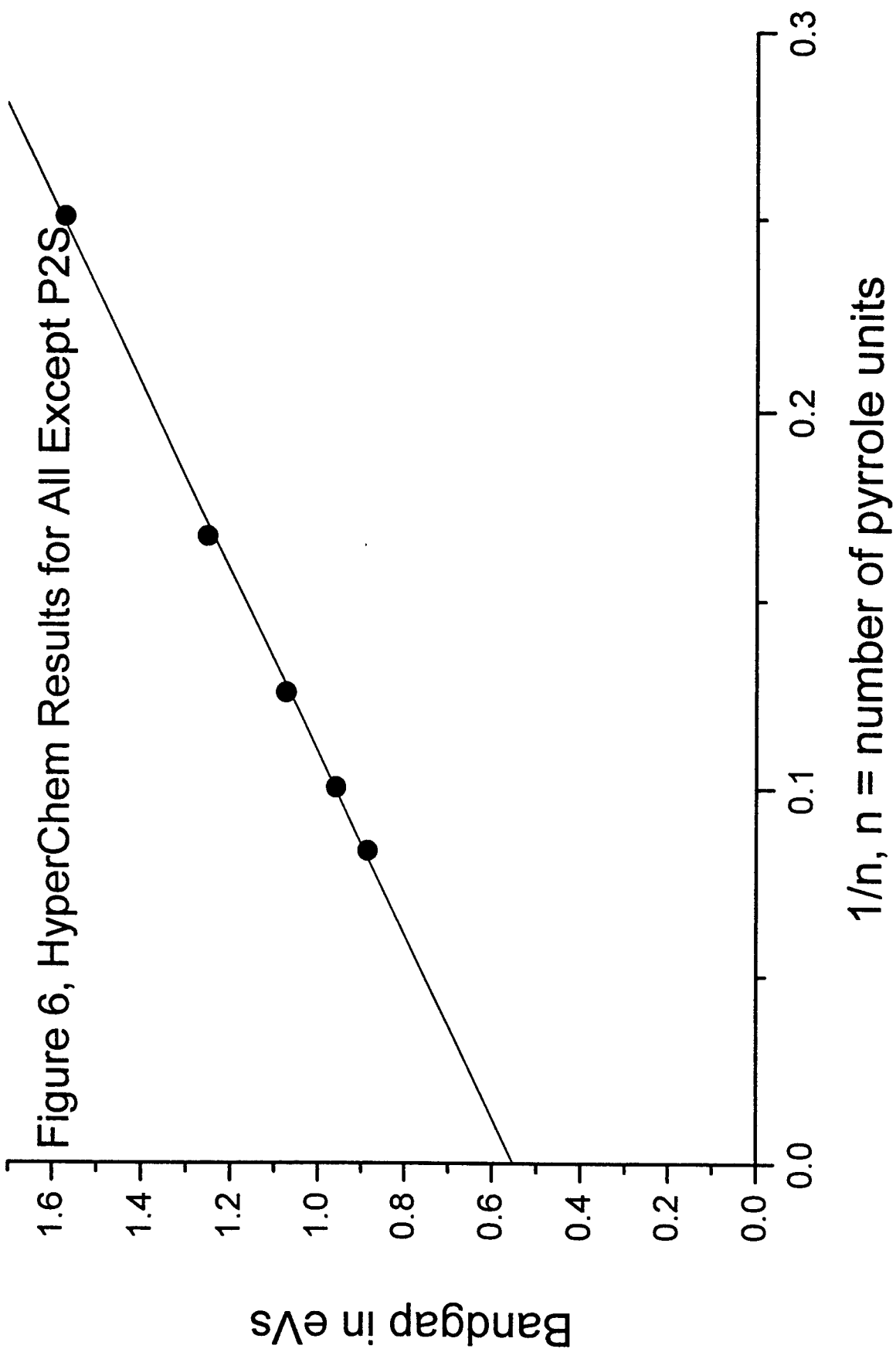


Table 3: Electron Affinities from GAMESS (in eVs)

Oligomer Designation	Electron Affinities
P2S	0.814
P4B	1.294
P6S	1.927
P8B	1.914
P10S	-
P12B	-

FINAL REPORT: CLASSIFICATION AND VISUALIZATION OF TISSUE IN MULTIPLE MODALITIES OF BRAIN MR

**Timothy S. Newman, Ph.D.
Department of Computer Science
University of Alabama in Huntsville
Huntsville, AL 35899**

**Final Report for Summer Research Extension Program in
extension of 1996 Wright Laboratory (WL/AARA) Summer Faculty Program**

**Sponsored by Air Force Office of Scientific Research
Bolling AFB, DC**

May 1998

**FINAL REPORT: CLASSIFICATION AND VISUALIZATION OF
TISSUE IN MULTIPLE MODALITIES OF BRAIN MR**

**Timothy S. Newman, Ph.D.
Department of Computer Science
University of Alabama in Huntsville
Huntsville, AL 35899**

Abstract

In this project, tissue classification and visualization in multiple modalities of magnetic resonance (MR) data of the brain was investigated. The project focused on enhancement of tissue segmentation and classification methods and exploration of effective visualization of anatomical structures. Much of the effort examined extracting and rendering ventricular and vascular structures. A number of visualization techniques were developed and applied to MR brain datasets. The techniques are described and sample final visualizations are exhibited in this report.

FINAL REPORT: CLASSIFICATION AND VISUALIZATION OF TISSUE IN MULTIPLE MODALITIES OF BRAIN MR

Timothy S. Newman, Ph.D.
Department of Computer Science
University of Alabama in Huntsville
Huntsville, AL 35899

1 Introduction

In this report, we describe our efforts to effectively visualize the brain's ventricular and vascular systems in multiple modalities of magnetic resonance imaging (MRI) data. These efforts are a continuation of the work that began at Wright Labs in Summer 1996 in collaboration with James Leonard. The investigation of visualization of the ventricular system involved four modalities of MR data, including magnetic resonance angiography (MRA), T1-weighted, T2-weighted, and proton density datasets. Visualization of the brain's vascular system was investigated using MRA data.

This report is organized as follows. First, the problem that was investigated is motivated. Second, some of the related projects from the literature are described. Next, an overview of the work that was conducted at Wright Lab in 1996 is presented and finally the extension of that work in this extension project is described and some results are exhibited.

2 Motivation and Background

The motivating factors for this research included enhancing capabilities to exploit pre-operative (and possibly intra-operative) image data in surgical planning and guidance. Providing anatomical and functional information to clinical staff can also aid in diagnostic assessment and treatment monitoring. A number of papers in the literature have described the value of high-resolution high level information (for example, visualization of the 3D boundaries of all relevant structures) in a region of interest for surgery planning. For example, the reader may wish to consider the recent papers [1, 30, 31]. For surgical and radiotherapy interventions deep in the brain, it is especially useful for the locations of cortical structures and blood vessels to be well-known [19, 30].

The goals of this research project were to study methods to enhance understanding of locations of various tissue and structures within the brain. The work has focused primarily on improving the segmentation of the brain's ventricles and the presentation of meaningful and efficient visualizations of the vascular system. The ventricular and vascular systems of the brain are two structures whose accurate visualization can aid greatly in diagnostic assessment, intervention planning and guidance, and evaluation of treatment.

Issues in the registration of datasets have previously been studied by several research teams. Several techniques have been presented for dataset registration. The traditional registration method involves the use of a stereotactic frame rigidly attached to a patient. Some of the projects reported in the literature have used external fiducial markers that are visible in the acquired data for registration. For example, Adams et al. [1], the Cleveland Clinic group [3], and Malison et al. [22] have used external markers to aid in registration. Frameless surgery is also possible. For example, Peters and colleagues [30] at the Montreal Neurological Institute have registered datasets by manually choosing landmark points on the surface of the head. They calculated the transformation that aligns the data points using least squares.

Hemler et al. [14] have registered MR and CT datasets by matching polygonized boundaries of the head. A number of researchers have attempted to register patient datasets with brain atlas data using deformation-based techniques. For example, Christensen et al. [5, 6] have used models of elastic solids and viscous fluids as motivation for their deformation method.

Methods to segment various structures of interest in tomographic data have been investigated by many research groups. Fishman et al. [8] have outlined the general classes of segmentation approaches in the medical imaging literature. Some of the segmentation methods for structures in the brain include the methods of Li et al. [21] who segment major tissue classes (white matter, gray matter, cerebro-spinal fluid, and potential tumors) in MR datasets using symmetry, morphological operations, and measures of segmented region size. Arata et al. [2] have used histogramming and K-means clustering to segment ventricles, sulci, and white matter lesions from proton density and T2-weighted data. Kapouleas [17] has extracted the outline of the brain and suspected Multiple Sclerosis lesions using local edge detection in proton density and T2-weighted data.

3 Overview of Efforts in Summer 1996

In the 1996 Summer Faculty Research Program, extraction of certain anatomical structures in head MR images was investigated. The goal of that effort was to first determine anatomical landmarks that were appropriate to allow the registration of multiple modalities of MR. Registration (that is, alignment) of MR datasets offers the potential of exploiting the capabilities of each modality in an integrated fashion. (For example, the vasculature tends to have the highest contrast with other tissue in MRA data, making MRA the best MR modality for studying the vasculature. In T2-weighted images, edema caused by tumor cell invasions is highly contrasted with other tissue [19]. If the relative orientation of the subject in each dataset is known, the ensemble of data can allow a wide variety of useful structures' relative positions to be better understood.) Several landmarks seemed to be visually apparent in the four modalities of MR data available for the project. These included the eyes, the lateral ventricles, and the longitudinal fissure of the brain. These landmarks could be used to compute the transformations that align the datasets.

After the appropriateness of these anatomical landmarks was determined, techniques were developed and/or applied to the task of determining landmark location in each image modality. The techniques will be briefly described here. A more complete description is available in the 1996 program final report [24] and in a 1997 conference paper [27].

An algorithm that the PI had previously developed for extraction of the kidney, kidney collecting system, kidney cysts, and other nearby organs in lower torso CT data [29] was applied to the problem of segmenting the position of the lateral ventricles in the four modalities of MR data. The method requires a human to select an initial seed in the structure of interest and uses multiple stages of volume-growing, morphological refinement, and locally adaptive histogramming to achieve a final segmentation.

For extraction of the longitudinal fissure of the brain, the method of Brummer [4] was applied. Brummer's method assumes that the longitudinal fissure can be well-modeled by a plane, and he proposed an efficient method to automatically determine that plane in MR data. During the 1996 summer program, we began implementation of the Brummer method, but were unable to complete its implementation before the end of the program.

Most of the effort during the summer involved development of a new algorithm to automatically locate the eyes in the four modalities of MR data available. The method developed exploits the spherical shape of the eye to achieve segmentation of the eyes. The method uses the Hough Transform circle detection scheme [20] to find circular regions in edge images formed from each slice. The set of circles is then clustered into sets. Finally, the method of Hall et al. [12] is used to determine the best-fitting sphere to each set, and the two best-fitting spheres which have radii within the allowable range for eyes are accepted as the position of the eyes. A detailed description of the eye extraction technique is available in [24] and [27].

4 Extension Program Effort

In the followon project to the 1996 summer effort, we improved the segmentation of the ventricular system and developed methods to improve classification of tissue representing blood vessels. In addition, we developed methods to improve final rendering and visualization of these structures. In this section, the methods will be briefly described and some results will be exhibited.

4.1 Segmentation and Visualization of Vasculature

The methods for segmenting and visualizing vasculature will be discussed first. The effort to localize vasculature and to present an effective visualization began with an investigation of the maximal intensity projection.

4.1.1 MIP

Maximal intensity projection (MIP) is one method to generate a direct volume rendering. Direct volume rendering can be contrasted with the traditional method of graphical rendering, the surface-based rendering. In surface-based rendering, the output image is constructed by applying a physically-motivated lighting model to surfaces in a scene. Volume-based rendering does not rely on the explicit determination of surface features and subsequent application of a lighting model to the surface features.

The MIP is a two dimensional representation of a three (or higher) dimensional dataset. Each point P_i in the MIP is formed from casting a ray that originates at the point and passes through the volume. The intensity, I_i , of P_i is set to the maximal intensity value of all the voxels intersected by the ray as it passes through the volume.

One example maximal intensity projection image from a brain MRA dataset is shown in Figure 1. In this image, the rays were projected perpendicular to the image plane and along one of the axes of the grid of MRA data. Because flow is captured as high intensity voxels in MRA data, blood flow tends to be one of the highest contrast tissues in MRA. The whitish trails in the image are the blood vessels.

MIP images typically have an X-ray-like appearance. While these images allow an observer to discern the general pattern of the blood vessels in the volume, the relative three-dimensional locations of the blood vessels is not apparent. However, we have experimented with two techniques that allow a viewer to have an improved perception of the relative locations of the blood vessels.

The first technique involves the generation of a series of MIP images from an image plane which is rotated in a circular orbit around the central axis of the volume. By performing MIP at regular intervals in this orbit and then



Figure 1: Sample Maximal Intensity Projection on brain MRA dataset.

“playing” the sequence of images as a cinematic loop, a viewer perceives the position and depth of many of the major blood vessels. We have placed one example animative rendering on the web. It can be viewed at the URL <http://www.cs.uah.edu/visgig/K12/images/animation.gif>.

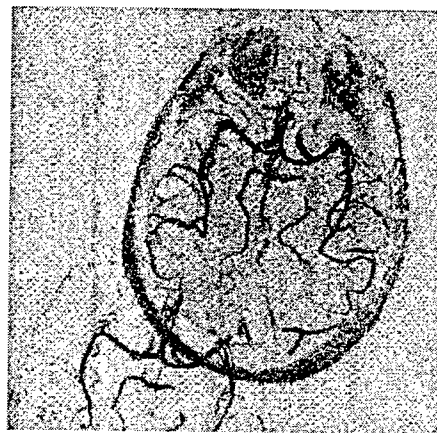


Figure 2: 3D rendering of volume of MIP points.

The second visualization technique involves generating a 3D volume using the MIP image. If the position of the maximal-valued voxel for each ray is recorded, a three-dimensional representation of the blood vessels can be created. Standard graphics depth-cueing can be applied to the realization of these points to allow a user to perceive the position of the points. In addition, the points can be viewed in three space using the 3D rendering capabilities of a visualization package. An example visualization is shown in Figure 2. The MIP volume is shown in the middle of the image and a projection of that data to a bounding plane is shown in the lower left corner of the image. The resultant image can also be viewed using a rainbow colormap to display the depths. A black and white rendering of the color image is shown in Figure 3. (The color image reveals much more information, but reproduction limitations restrict us to black and white copy in this report.) It is possible for clinical staff to exploratively interact with this data; the user can rotate and translate the data to view it from different directions and to thereby understand the spatial relationships of the structures.

Two sample visualizations of the vasculature are shown in Figures 4 and 5. The visualizations were computed on the MRA dataset of a healthy patient. The images displayed in these figures are black and white representations of color images that were synthesized in the research project. These visualizations show maximal intensity projections from three orthogonal directions that are projected onto the sides of the cube which bounds the data. In addition, a

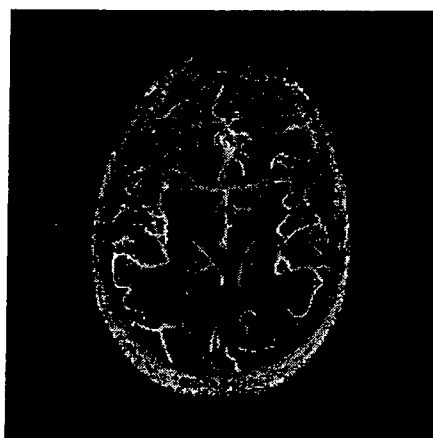


Figure 3: Black and white rendering of depth-colored MIP.

fourth volumetric projection from an oblique direction is shown as a shadow below the cube.

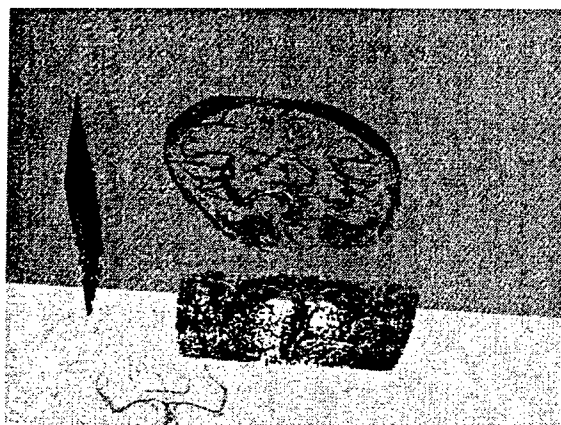


Figure 4: Visualization showing 3 MIPs projected to bounding cube.



Figure 5: Second visualization showing 3 MIPs projected to bounding cube.

We are currently investigating using a collection of MIPs for visualization. Those MIPs would be collected from viewpoints that are approximately equally distributed on the surface of a sphere that contains the volume and whose centroid is coincident with the centroid of the volume. We are investigating if a collection of many MIPs in MRA data begins to approach the appearance of an isosurface rendering on the data. We have collected isosurface renderings with fixed thresholds on the MRA data, but the isosurface renderings are not reproduced here. We want to determine how many MIP viewpoints are necessary to reasonably reproduce the position of nearly all of the vasculature and to determine how many viewpoints are needed before that data approaches the set of isosurface data.

The advantage of MIP over isosurface extraction is that the MIP method does not require a user to establish a threshold level, thus MIP may have the potential for more automatic extraction and visualization without the need to carefully select threshold parameters. We are continuing to investigate this potential.

4.1.2 Linking in Projection Images I - 2D distance

The first attempt at improving the MIP data to allow better visualization of vasculature and in turn more useful clinical application involved attempting to link some of the blood vessels that were disconnected in binarizations of MIP images. Due to partial volume effects and some in-plane flow, vasculature in the MRA data is not always uniformly "bright." As a result, there tends to be no fixed threshold that can satisfactorily reveal only vasculature, whether that threshold be applied as the isosurface level or as a threshold on the result of maximal intensity projection images. Consequently, vasculature tends to be disconnected in any 2D or 3D renderings of MRA data.

To connect the disconnected vessels in the processed data, we first used a linkage technique based on the well-known distance transform. Two other techniques for linkage were considered but only cursorily explored. These included morphological closing and volume growing in T1 data using seed points from the highest intensity voxels in the MIP on the MRA data. Morphological closing tended to produce unwanted effects as it closed gaps between vessels, although possibly subsequent processing could be developed to overcome such artifacts. Volume-growing in T1 data was not attempted because the amount of vasculature that is visually distinguishable in T1-weighted MR data appeared to be less than that of MRA; we believe that fewer vessels would be extracted in T1. However, it might be possible to improve the segmentation of a few key vessels by coupling T1 region-growing of blood vessels with the MRA-based methods.

Our linkage method used both pixel distance and intensity. The standard Distance Transform links segments based only on their distance. Thus, linear segments tend to result from Distance Transform-based linkage. By using the intensity of the pixels in the original MIP data, the path of linkage can be influenced when there is intensity evidence indicating the potential of a blood vessel within tissue. Our method weights both the distance and intensity to do the linkage.

The linking method first identifies isolated components. Components consisting of a single pixel are removed from consideration to avoid linking with isolated noise points. Next, for each component, the lowest cost path to any nearby component is found. Our method allows paths of distance at most 7 pixels.

Improved perception of vasculature is possible if the linked pixels are added to volumetric representations computed from single or multiple-direction MIP data.

Examples of the MIP projections from two directions are shown in Figures 6 (a) and (b). A binarizing thresholding of these projections are shown in (c) and (d). The bottom two images, Figures 6(e) and (f), show the improvement from applying the Distance Transform-based edge linkage. The number of isolated vessel regions is reduced in the edge-linked images.

4.1.3 Linking in Projection Images II - 3D distance

A three-dimensional linking algorithm based on a 3D distance transform was also investigated. The algorithm is a three dimensional analogue of the essentially 2D linking in individual MIPs. The paths of linkage are computed in the volume of data that is formed from the sparse 3D representation of a single (or of multiple) MIPs. That volume contains data only for voxels that were a maximal point on one of the projection rays of one (or more) MIPs. By linking disconnected components in such data, a true three dimensional representation of the vasculature can be found.

The method requires large memory spaces to execute fairly efficiently.

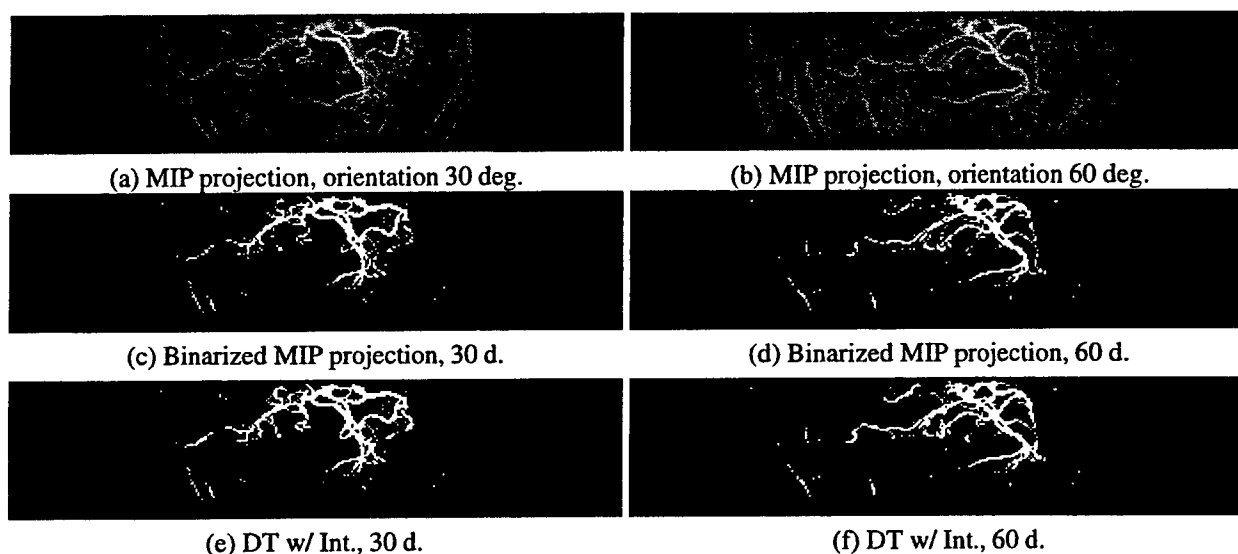


Figure 6: MIP Projections and Vascular Linkage.

In our experiments, we have thresholded the output from the MIP; low-intensity points from MIP are rejected before computing the volume linking.

Experiments are currently being conducted to ascertain the impact of the number of MIP projections used to form the initial volume. All of the MIP linkage experiments aim to overcome the limitations caused by selection of a threshold that removes much of the background “noise” found by the MIP method while also removing some of the vasculature. The exhibited images demonstrate the improvement in segmentation and visualization of vasculature using the linkage.

4.1.4 Linking Using Probabilistic Relaxation

The PI has recently developed a technique for segmentation of vehicles in synthetic aperture radar (SAR) [26]. That method involves a histogram-based initial segmentation followed by a series of refinement stages that link disconnected regions in an image. We attempted to apply this method to the individual 2D MIP data, both for segmentation and then for linking of the vessel segments. However, the initial segmentations were not good, and the final linkings seemed to be inferior to the other methods we tested.

Full details of the method were presented in [26]. The method computes a histogram for a slice image and attempts to find the division between the first and second mode. Hierarchical clustering is then used to group pixels that are members of the second mode. The clustering uses a weighting based on both pixel position and intensity. A relabeling step motivated by the probabilistic relaxation labeling is used to create a final segmentation. The relabeling tends to link disconnected components.

Additional investigation is required to be able to successfully apply this method to the vascular segmentation problem in MRA data. The MRA data tends to have three modes, including the background around the head, skin and fatty tissue, and everything else. The third mode is the one that includes the blood vessels, although the vessels are far out in the tail of the mode. Thus, histogram-based segmentations of vasculature in the MIP data may be difficult.

4.2 Ventricle Segmentation and Visualization

In the follow-on effort, the segmentation method used to segment the brain's ventricles was refined. The technique produces excellent segmentations of the lateral ventricles, although the smaller third and fourth ventricles are more difficult to segment. A journal paper describing the method has been submitted recently [28].

The method was originally applied to the segmentation of the kidneys, kidney collecting system, cysts and tumors on and adjacent to the kidneys, and other organs in lower torso CT data.

There are two major stages to the method. The first stage involves volume-growing from a human-supplied seed. Thus, the users of the method chooses a voxel that is within a lateral ventricle. A volume is grown out from the seed point using connectedness and prior knowledge about intensity distribution of CSF in MR data. The first stage produces a coarse segmentation of the lateral ventricles. The second stage of processing entails a morphological operation followed by locally adaptive histogramming. The locally adaptive histogramming refines the initial segmentation. A histogram is computed in a region of each slice about the original segmented volume. This second stage of processing was the major refinement of the segmentation method during the extension project.

The second stage of processing involves first dilating the volume-grown structure based on the size of the structure. For each slice, a histogram is computed for the dilated region. In the slice S_i of the initial seed, the histogram mode M_i that contains the seed is automatically identified. Then, for the slices S_{i+1} and S_{i-1} adjacent to that slice, the peak of mode M_i is taken as a point that defines the mode of interest in those slices. This process is repeated for successive slices; the peak of mode M_j in slice S_j is taken as a point within the mode of interest, M_{j+1} of slice S_{j+1} . This process defines the mode of interest in each slice, and the pixels that fall within that mode are taken as the final segmentation of the structure of interest within the node.

An example showing an initial image and the final segmentation of one ventricle is shown in Figures 7 and 8, respectively. Two shaded three-dimensional renderings of that extracted ventricle are shown in Figure 9

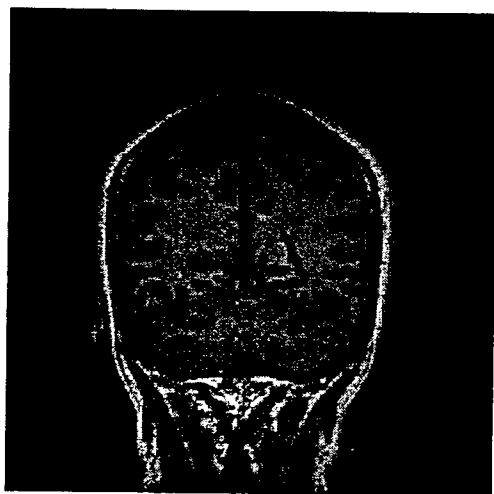


Figure 7: T1-weighted slice containing ventricles.

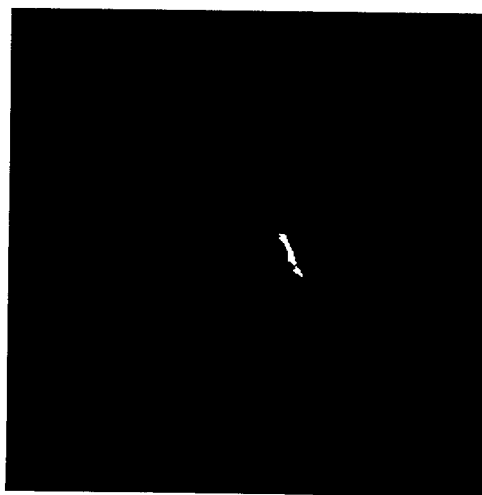


Figure 8: The portion of extracted left lateral ventricle in the same slice as at left.



Figure 9: Shaded 3D Renderings of Extracted Ventricle.

4.3 Miscellaneous Additional Action Items

The PI also was able to observe an image-guided pedicle screw surgery at the Cleveland Clinic during the project. Additionally, dicom-format images from a patient with artero-venous malformation were obtained from the Cleveland Clinic. These images will be used in future investigations and to validate the work.

5 Conclusion and Future Work

This document has presented a summary of the research that was conducted in our 1996 Summer Research Extension Program. Methods for the improved classification and visualization of tissue representing cerebro spinal fluid (CSF; in particular, the ventricular system) and the vascular system of the brain have been developed. These techniques can be helpful for diagnostic assessment, intervention planning, and treatment assessment. Clinical staff can better understand relative anatomical location of structures of interest through our techniques.

5.1 Future Efforts

Future research efforts will extend the methods that we have developed. We are currently exploring the registration of MRA datasets collected at different times. We are investigating methods to extract landmark points and curves from the vascular tree that are suitable for registration. In addition, we are developing techniques to automate the selection of the seed point for the ventricle segmentation.

To date, our work in extracting the third and fourth ventricles has been cursory. These smaller structures do not seem to be as clearly distinguished in all four data modalities as do the larger lateral ventricles. Future work is needed to more robustly classify those structures. In addition, only cursory attention was focused on distinguishing between gray and white matter. Preliminary work suggests that histogram-based techniques in T2 data may allow coarse segmentation of these tissue types. Those coarse results are perhaps validated by the work of Hornak and Fletcher [16], who have previously reported success in segmenting gray and white matter using multidimensional histograms.

5.2 Demonstrable Achievements

One journal paper [28] has recently been submitted that is based partly on the work described in the extension project. That paper is currently under review. An additional journal submission is planned to present the method for eye extraction. A conference paper will be submitted soon that discusses our vascular segmentation and visualization method. Finally, one conference paper [27] has been presented that discusses the early stages of our work in anatomical structure classification in the brain. That paper seemed to have a good audience reception at the conference.

The project has produced some progress toward the goals of allowing more clear visualization of anatomical structures in the brain. We hope that future work will be possible to improve the methods further.

Acknowledgments

The Principal Investigator benefitted from many fruitful discussions with James Leonard during the course of the summer faculty program and the extension project. In addition, discussions with the PI's students, especially with Jonathan Bush, Vicky He, Shiva Nadarajah, and Kit-Cheng Ng were helpful. Some of the results presented in this report are the result of experiments performed by and/or algorithms implemented by these students under the PI's direction.

References

- [1] Ludwig Adams, Achim Knepper, Dietrich Meyer-Ebrecht, Rainer, Rüger, and Willem van der Brug, "An Optical Navigator for Brain Surgery," *IEEE Computer*, Vol. 29, No. 1, 1996, pp. 48-54.
- [2] Louis K. Arata, Atam P. Dhawan, Joseph P. Broderick, Mary F. Gaskil-Shipley, Alejandro V. Levy, and Nora D. Volkow, "Three-Dimensional Anatomical Model-Based Segmentation of MR Brain Images Through Principal Axes Registration," *IEEE Transactions on Biomedical Engineering*, Vol. 42, No. 11, Nov. 1995, pp. 1069-1078.
- [3] G. H. Barnett, D. W. Kormos, C. P. Steiner, D. Piraino, J. Weisenberger, F. Hajjar, C. Wood, and J. McNally, "Frameless Stereotaxy Using a Sonic Digitizing Wand: Development and Adaptation to the Picker ViStar Medical Imaging System," in *Interactive Image-Guided Neurosurgery*, edited by R. J. Maciunas, American Association of Neurological Surgeons: Park Ridge, IL, 1993, pp. 113-119.
- [4] Marijn E. Brummer, "Hough Transform Detection of the Longitudinal Fissure in Tomographic Head Images," *IEEE Transactions on Medical Imaging*, Vol. 10, No. 1, March 1991, pp. 74-81.
- [5] Gary E. Christensen, Richard D. Rabbit, and Michael I. Miller, "3D Brain Mapping using a Deformable Neuroanatomy," *Physics in Medicine and Biology*, Vol. 39, 1994, pp. 609-618.
- [6] Gary E. Christensen, Michael I. Miller, Michael W. Vannier, and Ulf Grenander, "Individualizing Neuroanatomical Atlases using a Massively Parallel Computer," *IEEE Computer*, Vol. 29, No. 1, 1996, pp. 32-38.
- [7] Richard O. Duda and Peter E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons: New York, 1973.
- [8] E. Fishman, B. Kuszyk, D. Heath, and L. Gao, "Surgical Planning for Liver Resection," *IEEE Computer*, Vol. 29, No. 1, 1996, pp. 64-72.
- [9] J. Foley, A. van Dam, S. Feiner, and J. Hughes, *Computer Graphics: Principles and Practice, Second Edition in C*, Addison-Wesley: New York, 1996.

- [10] Jim C. Gee, Martin Reivich, and Ruzena Bajcsy, "Elastically Deforming 3D Atlas to Match Anatomical Brain Images," *Journal of Computer-Assisted Tomography*, Vol. 17, No. 2, 1993, pp. 225-236.
- [11] W. Eric L. Grimson, G. J. Ettinger, S. J. White, T. Lozano-Pérez, W. M. Wells, and R. Kikinis, "An Automatic Registration Method for Frameless Stereotaxy, Image Guided Surgery, and Enhanced Reality Visualization," *IEEE Transactions on Medical Imaging*, Vol. 15, No. 2, 1996, pp. 129-140.
- [12] Ernest L. Hall, James B. K. Tio, Charles A. McPherson, and Firooz A. Sadjadi, "Measuring Curved Surfaces for Robot Vision," *IEEE Computer*, Vol. 15, 1982, pp. 42-54.
- [13] Martial Hebert and Jean Ponce, "A New Method for Segmenting 3-D Scenes Into Primitives," *Proceedings of the Sixth International Conference on Pattern Recognition*, 1982, pp. 836-838.
- [14] Paul Hemler, Thilaka Sumanaweera, Petra van den Elsen, Sandy Napel, and John Adler, "Quantified Registration Error Versus the Accuracy of Registered Surfaces for a Multimodality Surface-Based Registration System," *Proc., SPIE International Symposium on Medical Imaging 96*, Newport Beach, CA, Feb. 1996.
- [15] K. H. Höhne, B. Pflessner, A. Pommert, M. Riemer, T. Schiemann, R. Schubert, and U. Tiede, "A 'Virtual Body' Model for Surgical Education and Rehearsal," *IEEE Computer*, Vol. 29, No. 1, 1996, pp. 25-31.
- [16] Joseph P. Hornak and Lynn M. Fletcher, "Multispectral Image Segmentation in Magnetic Resonance Imaging," in *Digital Image Processing Methods*, ed. by Edward R. Dougherty, Marcel Dekker: New York, 1994.
- [17] Ioannis Kapouleas, "Segmentation and Feature Extraction for Magnetic Resonance Brain Image Analysis," *Proc., Tenth International Conference on Pattern Recognition*, Atlantic City, NJ, June 1990, pp I-583 to I-590.
- [18] A. Kaufman, "Trends in Volume Visualization and Volume Graphics," in *Scientific Visualization: Advances and Challenges*, ed. by L. Rosenblum et al., Academic Press: San Diego, 1994.
- [19] P. J. Kelly, *Tumor Stereotaxis*, W. B. Saunders: Philadelphia, 1991.
- [20] Carolyn Kimme, Dana Ballard, and Jack Sklansky, "Finding Circles by an Array of Accumulators," *Communications of the ACM*, Vol. 18, No. 2, Feb. 1975, pp. 120-122.
- [21] Chunlin Li, Lawrence O. Hall, and Dmitry B. Goldgof, "Knowledge-Based Classification and Tissue Labeling of Magnetic Resonance Images of Human Brain," *Proc., SPIE Conf. on Biomedical Image Processing and Biomedical Visualization*, SPIE Vol. 1905, San Jose, 1993, pp. 554-565.
- [22] Robert T. Malison, E. G. Miller, R. Greene, G. McCarthy, D. S. Charney, and R. B. Innis, "Computer-Assisted Coregistration of Multislice SPECT and MR Brain Images by Fixed External Fiducials," *Journal of Computer-Assisted Tomography*, Vol. 17, No. 6, 1993, pp. 952-960.
- [23] Timothy S. Newman, "Issues and Directions for Medical Applications of Computer Vision and Data Visualization," *Proceedings of the IEEE Workshop on Visualization and Machine Vision*, Seattle, June 1994, p. 106.
- [24] Timothy S. Newman, "Final Report of a Summer Faculty Project for Anatomical Feature Extraction for Registration of Multiple Modalities of Brain MR," *Proc., 1996 Air Force Summer Faculty Research Program*, pp. 44-1 to 44-20.
- [25] Timothy S. Newman, P. J. Flynn, and A. K. Jain, "Model-Based Classification of Quadric Surfaces," *Computer Vision, Graphics, and Image Processing: Image Understanding*, Vol. 58, No. 2, 1993, pp. 235-249.
- [26] Timothy S. Newman, Jinsoo Lee, and Scott R. Vechinski, "Target Extraction using Hierarchical Clustering with Refinement by Probabilistic Relaxation Labeling," *Proc., Automatic Target Recognition VIII Conf.*, (proceedings in press), Orlando, April 1998.
- [27] Timothy S. Newman and James D. Leonard, "Anatomical Feature Extraction for Registration of Multiple Modalities of Brain MR," *Proc., Biomedical Sensing, Imaging, and Tracking Technologies II Conf.*, San Jose, Feb. 1997, pp. 205-213.

- [28] Timothy S. Newman and Ning Tang, "Multistage Segmentation Procedure for Extraction of Anatomical Structures in Volume Data," submitted to *Pattern Recognition*, 1998.
- [29] Timothy S. Newman, Ning Tang, Stephen Bacharach, and Peter Choyke, "A Volumetric Segmentation Technique for Diagnosis and Surgical Planning in Lower Torso CT Images," *Proc., Thirteenth International Conference on Pattern Recognition*, Vienna, August 1996, pp. C-553 to C-557.
- [30] Terry Peters, Bruce Davey, Patric Munger, Roch Comeau, Alan Evans, and André Olivier, "Three-Dimensional Multimodal Image-Guidance for Neurosurgery," *IEEE Transactions on Medical Imaging*, Vol. 15, No. 2, 1996, pp. 121-128.
- [31] Joseph R. Smith and Herman F. Flanigin, "Stereotactic Neurosurgery," in *Essentials of Neurosurgery: A Guide to Clinical Practice*, edited by Marshall B. Allen and Ross Miller, McGraw-Hill: New York, 1995, pp. 463-478.

FPGA IMPLEMENTATION OF THE XPATCH RAY TRACER

Mohammed Y. Niamat
Associate Professor
Department of Engineering Technology

The University of Toledo
2801 W. Bancroft Street
Toledo, OH 43606

Final Report for:
Summer Research Extension Project
Wright Laboratory Avionics Directorate (WL/AASH)
Wright Patterson Air Force Base
Dayton, Ohio

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Wright Patterson Air Force Base
Dayton, Ohio

December 1997

FPGA IMPLEMENTATION OF THE XPATCH RAY TRACER

Mohammed Y. Niamat
Associate Professor
Department of Engineering Technology
The University of Toledo

Abstract

The ray tracing algorithm, often used in computer graphics for creating a 2-D picture of a 3-D world, is computationally intensive. Many attempts have been made in the past to reduce the processing time taken by this algorithm. Related work can be found in references [1] to [5]. In this research project, a systolic array architecture and its functional modules are developed for implementing the algorithm on Field Programmable Gate Array(s). Since FPGA(s) are programmable in real-time, this method has a definite advantage over conventional ones. It is hoped that the work will lead to the successful implementation of the algorithm on both the CHAMP and DRASTIC field programmable boards. Since the ray tracing algorithm is also used in the Xpatch software used for accurate electromagnetic scattering predictions, this work will also contribute toward reducing the processing time of the Xpatch ray tracer.

FPGA Implementation of the XPATCH Ray Tracer

Mohammed Y. Niamat

Introduction: The Ray Tracing Problem

For a given starting point $R(o)$, and a starting direction $R(d)$, the ray tracing problem is to determine if the ray defined by

$$R(t) = R(o) + R(d) * t, \quad t > 0$$

hits the target. If it does, the next question that needs to be answered is “where is the hit-point ? ” To accomplish these before ray tracing a target, a “tree” of the target is built. The tree contains information on how the facets are distributed in the space and is used to identify those facets that are likely to be hit by a ray. In *xpatchf*, the tree is based on an algorithm called ‘binary space partitioning’ (BSP).

In order to speed the computations for ray-object intersection, the object is often enclosed in a bounding volume. The bounding volume can be of different shapes. Common bounding volumes being spheres, and rectangular boxes. Kay and Kajiya [], used the concept of slabs to form rectangular boxes as bounding volumes. A slab is simply the space between two parallel planes. The intersection of a set of slabs defines the bounding volume. The method relies on intersection of each pair of slabs by the ray, keeping track of the *near* and the *far* intersection distances. If the largest near value is greater than the smallest far value, then the ray misses the bounding volume; otherwise, it hits.

The Algorithm

One of the simplest finite bounding volumes is the intersection of two parallel planes each aligned so that their normals are in the same direction as the X, Y, and Z axes. The following algorithm determines whether the bounding box is hit by the ray, or not (miss).

Define the orthogonal box by two coordinates:

$$\begin{aligned} \text{box's minimum extent} &= B(0) = [BBP(0,0) \quad BBP(0,1) \quad BBP(0,2)] \\ \text{box's maximum extent} &= B(1) = [BBP(1,0) \quad BBP(1,1) \quad BBP(1,2)] \end{aligned}$$

The origin and direction of the ray is defined by:

$$\begin{aligned} R(\text{origin}) &= R(0) = [rs(0) \quad rs(1) \quad rs(2)] \\ R(\text{direction}) &= R(d) = [rd(0) \quad rd(1) \quad rd(2)] \end{aligned}$$

The ray can then be defined as the set of points on a line described by the following equation:

$$R(t) = R(0) + R(d) * t, \text{ where } t > 0$$

The above equation can be broken up into its three axes components as follows:

$$\begin{aligned} \text{Along X- Axis:} & \quad X(t) = rs(0) + rd(0) * t \\ \text{Along Y- Axis:} & \quad Y(t) = rs(1) + rd(1) * t \\ \text{Along Z- Axis:} & \quad Z(t) = rs(2) + rd(2) * t \end{aligned}$$

In the above equations, 0, 1, and 2 represent the X, Y, and Z axes, respectively.

Since $X(t)$, $Y(t)$, and $Z(t)$ satisfy all points on the X, Y, and Z components of the ray, they must also satisfy the points on the minimum and maximum extents of the box. Therefore,

$$\begin{aligned} \text{Along X- Axis:} & \quad \text{BBP}(0,0) = rs(0) + rd(0) * t \\ & \quad \text{BBP}(1,0) = rs(0) + rd(0) * t \end{aligned}$$

$$\begin{aligned} \text{Along Y- Axis:} & \quad \text{BBP}(0,1) = rs(1) + rd(1) * t \\ & \quad \text{BBP}(1,1) = rs(1) + rd(1) * t \end{aligned}$$

$$\begin{aligned} \text{Along Z-axis:} & \quad \text{BBP}(0,2) = rs(2) + rd(2) * t \\ & \quad \text{BBP}(1,2) = rs(2) + rd(2) * t \end{aligned}$$

From the above equations, 't' can be computed for each of the above six equations. For example, for the X-axis:

$$\begin{aligned} t1 &= [\text{BBP}(0,0) - rs(0)] / rd(0), \text{ for the minimum extent of the box, and} \\ t2 &= [\text{BBP}(1,0) - rs(0)] / rd(0), \text{ for the maximum extent of the box.} \end{aligned}$$

The ray algorithm for the intersection of the box , can be described as follows:

Set *tnear* = minus infinity, and *tfar* = plus infinity (i.e., arbitrarily large values).

For each pair of planes associated with X, Y, and Z (described here for the set of X planes) :

If the direction $rd(0)$ is equal to zero, then the ray is parallel to the planes,
so:

If the origin is not between the slabs, i.e., $rd(0) < \text{BBP}(0,0)$, or
 $rd(0) > \text{BBP}(1,0)$, then the ray misses the box.

Else, if the ray is not parallel to the planes, then

begin:

Calculate intersection distances of planes:

$$t1 = [BBP(0,0) - rs(0)] / rd(0)$$

$$t2 = [BBP(1,0) - rs(0)] / rd(0)$$

If $t1 > t2$, swap $t1$ and $t2$

If $t1 > t_{near}$, set $t1 = t_{near}$

If $t2 < t_{far}$, set $t2 = t_{far}$.

If $t_{near} > t_{far}$, box is missed

If $t_{far} < 0$, box is behind ray, so box is missed

end.

end of for loop.

If the box is hit, the intersection distance is equal to t_{near} and the ray's exit point is t_{far} . Figure 1(a) illustrates the case where the ray hits the box. Notice that t_{near} is less than t_{far} for this case. In Figure 1(b), t_{near} is greater than t_{far} , and therefore the ray misses the box.

The various symbols used in calculating the ray/box intersection are tabulated below.

Symbols	Meaning
BBP(0, 0)	Minimum extent of the Bounding Box Parent in the x-direction
BBP(1, 0)	Maximum extent of the Bounding Box Parent in the x-direction
BBP(0, 1)	Minimum extent of the Bounding Box Parent in the y-direction
BBP(1, 1)	Maximum extent of the Bounding Box Parent in the y-direction
BBP(0,2)	Minimum extent of the Bounding Box Parent in the z-direction
BBP(1,2)	Maximum extent of the Bounding Box Parent in the z-direction
rs(0)	X-component of the origin (starting point) of the ray
rs(1)	Y-component of the origin (starting point) of the ray
rs(2)	Z-component of the origin (starting point) of the ray
rd(0)	X-component of the directional vector of the ray
rd(1)	Y-component of the directional vector of the ray
rd(2)	Z-component of the directional vector of the ray
t_{near}	Distance of the nearest ray/box intersection point from the origin of ray
t_{far}	Distance of the farthest ray/box intersection point from the origin of ray

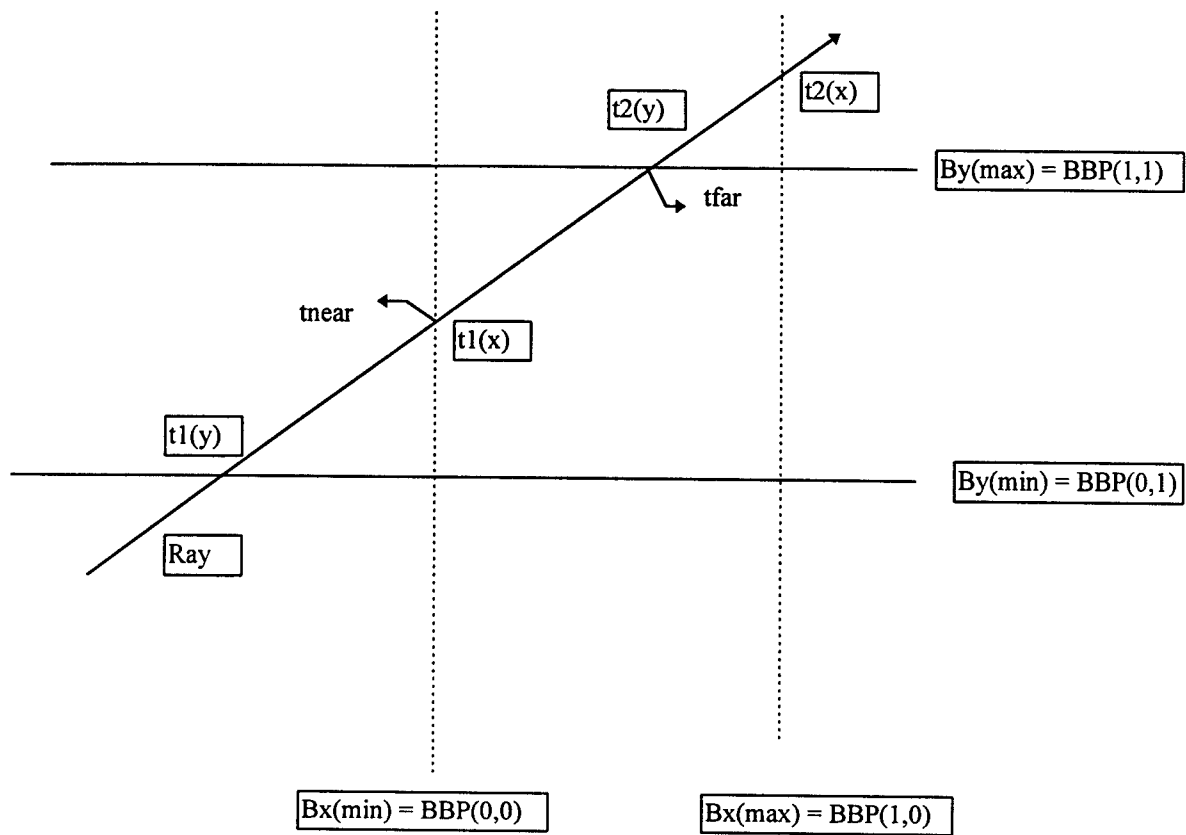


Figure 1 (a) : Ray Intersects Box

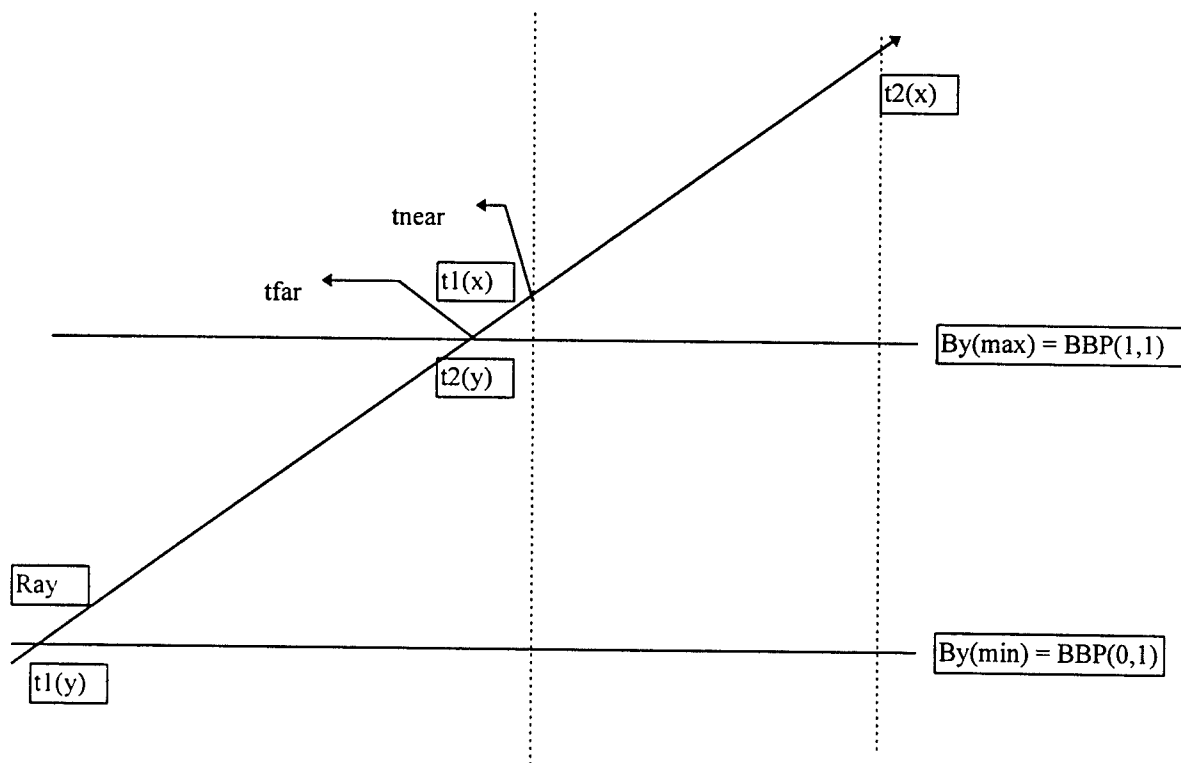


Figure 1 (b) : Ray Misses Box

The following numerical example illustrates the ray/box intersection problem.

Example

Given a ray with origin [0 4 2] and direction [0.218 - 0.436 0.873] and a box with corners:

$$B(\text{minimum}) = [-1 \ 2 \ 1]$$

$$B(\text{maximum}) = [3 \ 3 \ 3]$$

find if the ray hits the box.

Here, $rs(0) = 0$, $rs(1) = 4$, and $rs(2) = 2$;
 $rd(0) = 0.218$, $rd(1) = -0.436$, and $rd(2) = 0.873$;
 $BBP(0,0) = -1$, $BBP(0,1) = 2$, and $BBP(0,2) = 1$,
 $BBP(1,0) = 3$, $BBP(1,1) = 3$, and $BBP(1,2) = 3$.

The algorithm begins by looking at the X slab, defined by $BBP(0,0)$ and $BBP(1,0)$. The distances $t1$ and $t2$ are computed as:

$$\begin{aligned} t1 &= [BBP(0,0) - rs(0)] / rd(0) \\ &= [-1 - 0] / 0.218 = -4.59 \end{aligned}$$

$$\begin{aligned} t2 &= [BBP(1,0) - rs(0)] / rd(0) \\ &= [3 - 0] / 0.218 = 13.8 \end{aligned}$$

After this stage, $t1$ is set to $tnear$, and $t2$ to $t2$. Therefore $tnear$ becomes -4.59, and $t2$ 13.8. Neither $tnear > t2$ nor $t2 < 0$, so the Y slab is examined. For the Y slab:

$$\begin{aligned} t1 &= [BBP(0,1) - rs(1)] / rd(1) \\ &= [2 - 4] / -0.436 = 4.59 \end{aligned}$$

$$\begin{aligned} t2 &= [BBP(1,1) - rs(1)] / rd(1) \\ &= [3 - 4] / -0.436 = 2.29 \end{aligned}$$

Since $t1 > t2$, swap these values. The new $tnear = 2.29$, and the new $t2 = 4.59$. Next, we check for the Z slab.

$$\begin{aligned} t1 &= [BBP(0,2) - rs(2)] / rd(2) \\ &= [1 - 2] / 0.873 = -1.15 \end{aligned}$$

$$\begin{aligned} t2 &= [BBP(1,2) - rs(2)] / rd(2) \\ &= [3 - 2] / 0.873 = 1.15. \end{aligned}$$

At this stage, t_{near} is not updated, and remains 2.29. However, t_{far} is updated to 1.15. Since $t_{near} > t_{far}$, the ray misses the box.

Design Methodology

The following design steps are outlined to implement the ray tracing algorithm on the FPGAs:

1. Parallelize the algorithm.
2. Develop systolic arrays for the algorithm.
3. Synchronize the data items maintaining nearest neighbor connection.
4. Check theoretical output and obtain performance parameters (pipelining period, etc.).
5. Develop individual modules used in the overall design.
6. Interface the various modules to obtain the complete design.
7. Simulate and verify design.
8. Map design on Xilinx FPGA(s) using Xilinx's M1 Foundation and other CAD tools.

Proposed Architecture

The proposed systolic array architecture for finding the ray/box intersection is shown in Figure 2. It consists of nine P.E s (processing elements) connected in a systolic-type configuration. Each P.E. contains a subtractor and a multiplier. The first row computes the parameters t_1 and t_2 for the X- component, as defined earlier. The second and third rows compute parameters t_1 and t_2 for the Y and Z axes. The inputs $rs(0)$, $rs(1)$, and $rs(2)$, which define the starting point or origin of the ray, are fed from the left-hand sides of row 1, row 2, and row 3, respectively. The inverses for each of the directional data items, namely, $rd(0)$, $rd(1)$, and $rd(2)$ are first computed (not shown in figure). The results are then fed from the right hand side of the array as shown in the figure. The data items which define the minimum and maximum extent of the bounding box -- $BBP(0,0)$, $BBP(1,0)$, $BBP(0,1)$, $BBP(1,1)$, $BBP(0,2)$, and $BBP(1,2)$ are fed from the top of the array. The outputs t_1 and t_2 , for each of the axis, are obtained from the bottom of the cells as shown in the figure. The outputs t_1 for each of the axis are fed into a comparator circuit. The maximum of the three t_1 parameters is determined. This value is designated as t_{near} . In a similar fashion, the minimum of the three t_2 parameters is determined and designated as t_{far} . If t_{near} is greater than t_{far} , the ray misses the bounding volume box, otherwise, the ray intersects with the box and a hit is registered.

In a real target recognition application, it is expected that a considerably large number of rays will be fired in parallel at the target. This number could be as high as 25 million. In order to increase the speed of computations for a large number of rays, the architecture shown in Figure 2 can be duplicated. This will reduce the time for processing the rays to half. However, the number of processing elements will double to 18. If the architecture is replicated one more time, the number of cells would increase to 27. If computation of the 'plandist' parameter is neglected in the architecture, the number of cells used would be 6, 12, and 18 for the three cases under discussion.

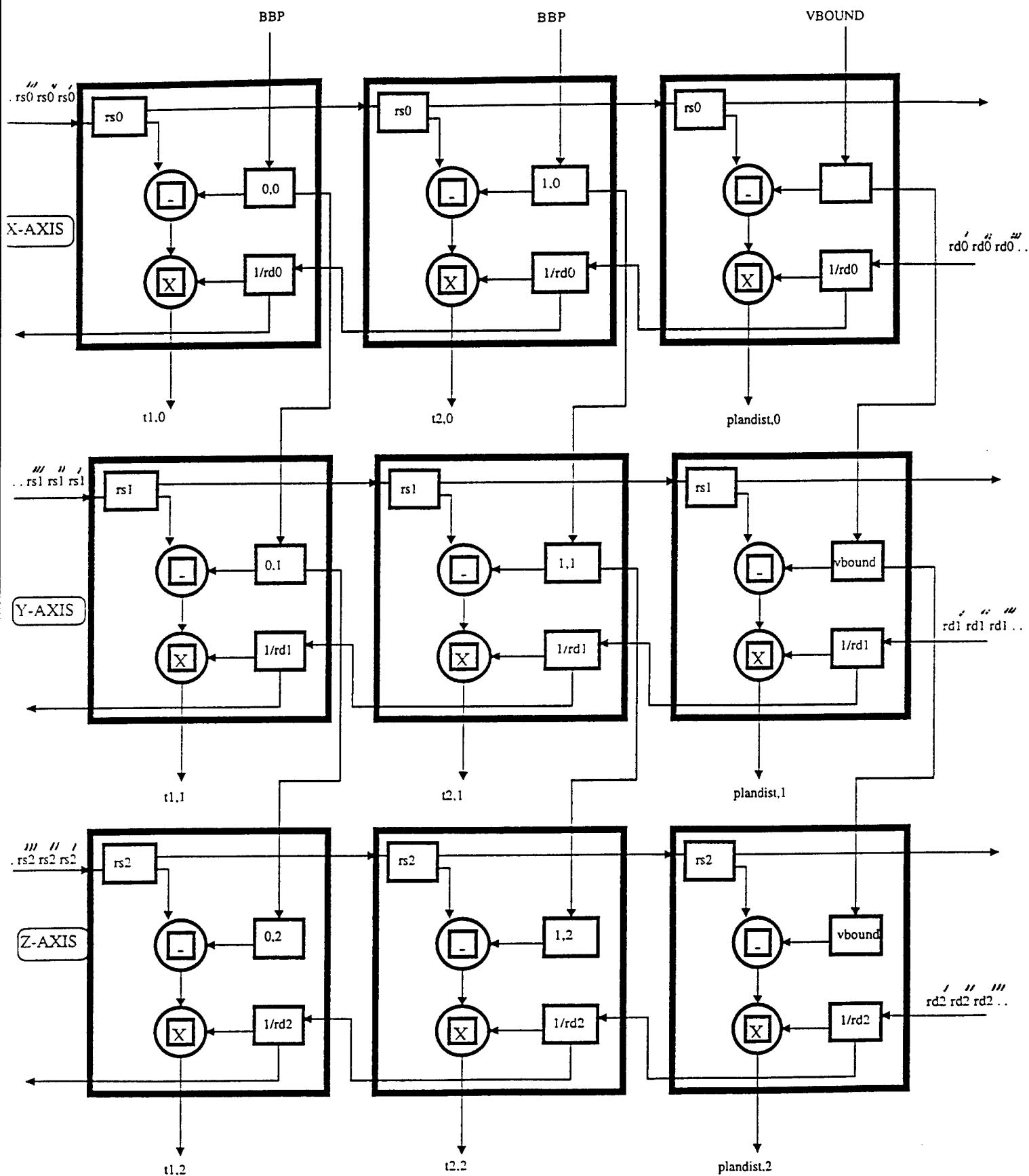


Figure 2 (a): Systolic Array, Time Snapshot 1

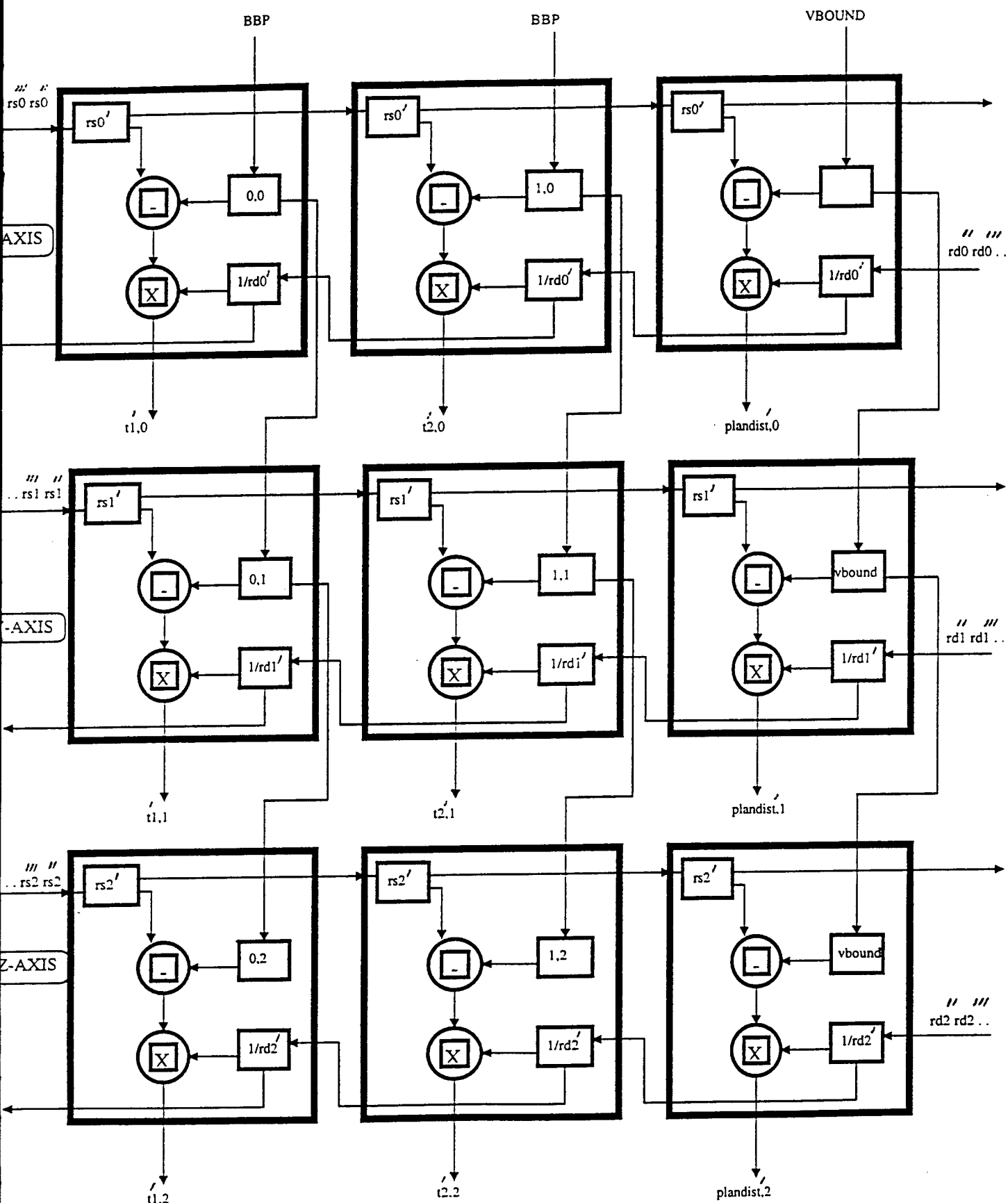


Figure 2 (b): Systolic Array, Time Snapshot 2

In order to implement the proposed systolic architecture, modules for the following components were developed and tested.

1. Carry Look-Ahead Adder
2. Wallace Tree Multiplier
3. Parallel in- Parallel Out Shift Right Register
4. Parallel in- Parallel Out Shift Left Register
5. Two's Complementer
6. Adder

1. Module: Carry Look -Ahead Adder

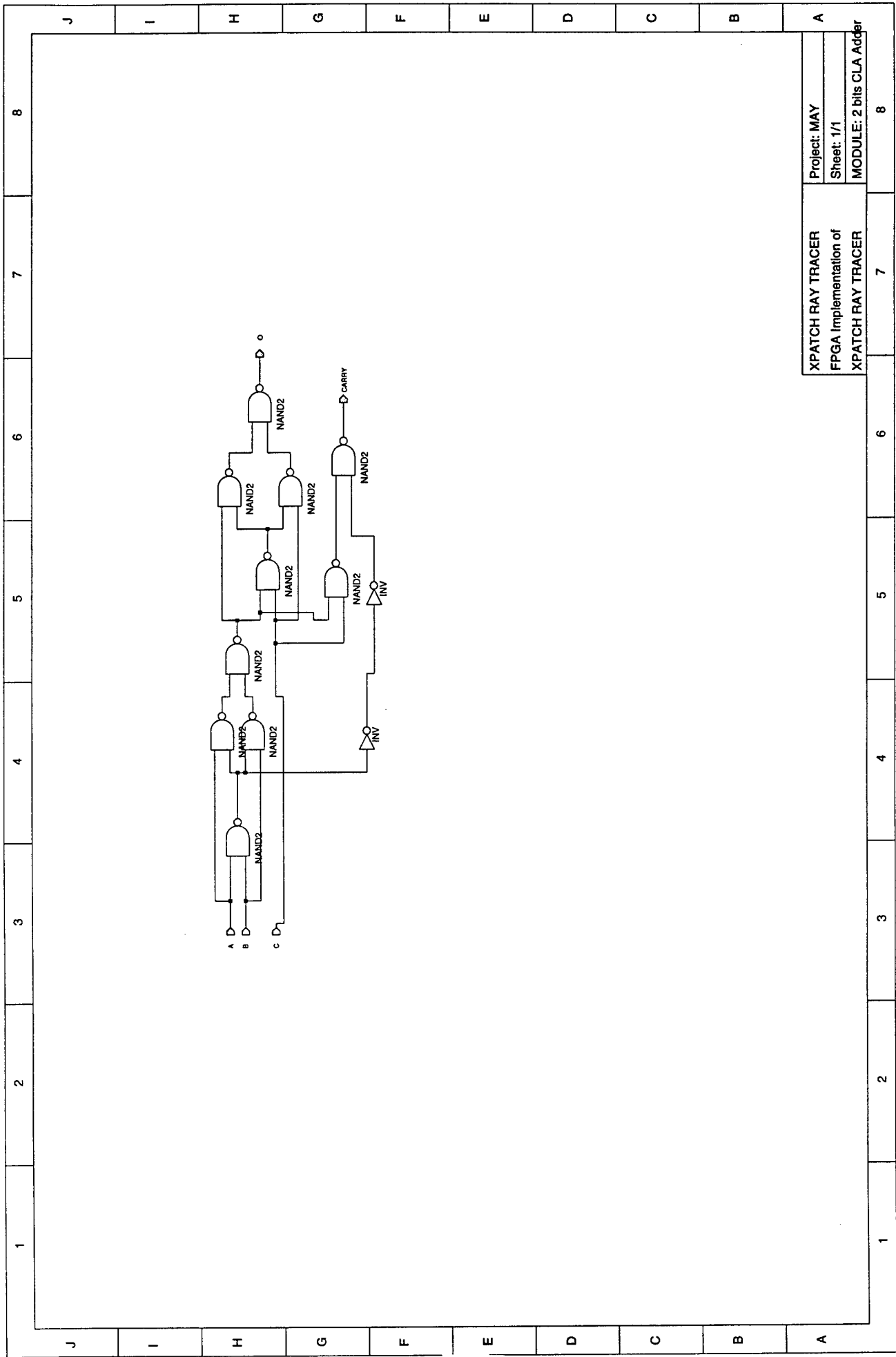
The Carry Look-Ahead module developed for this project is shown in Figure 3. It is designated as CLA2 and functions as a 2-bit Carry Look-Ahead adder . The components of the CLA2 include ten NAND and two Inverter gates. The following table summarizes the functions of the various input and output parameters of the CLA2.

Input Parameter	Function Description
A	Augend
B	Addend
C	Carry-in

Output Parameter	Function Description
O	Sum
Carry	Carry- out

SIMULATION RESULTS:

The simulation results for the carry-look-ahead adder are shown in Figure 4, and are found to be functionally correct. For example, it can be seen from the results in Figure 4 that for the time range 100ns to 200ns, if signal A = high, B = low, and C = high, then the signal 'O' for sum' is low and signal for 'carry' is high, as expected.



XPATCH RAY TRACER
 FPGA Implementation of
 XPATCH RAY TRACER

Project: MAY
 Sheet: 1/1
 MODULE: 2 bits CLA Adder

Figure 3: 2 bit CLA Adder

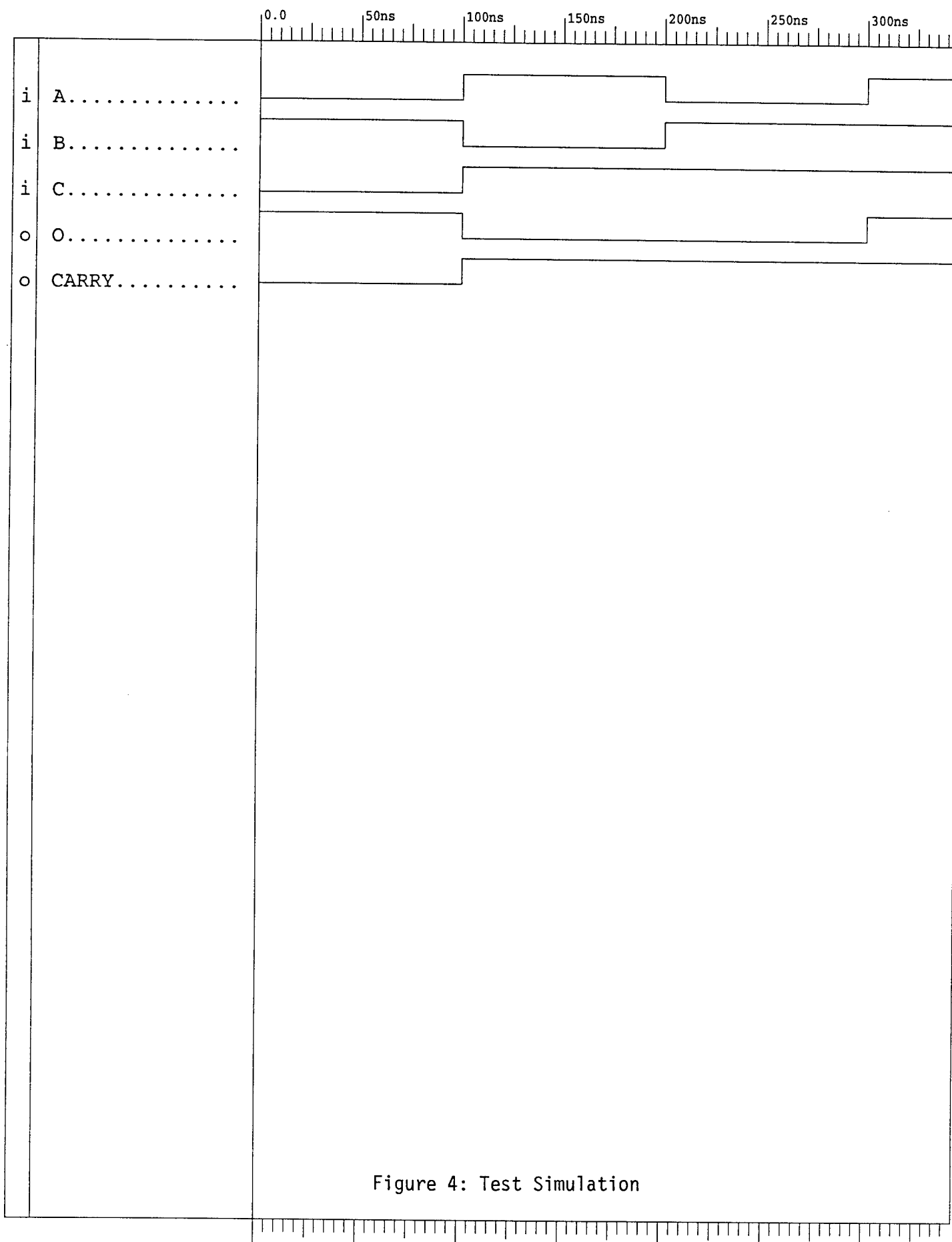


Figure 4: Test Simulation

2. Module: Wallace Tree Multiplier

The Wallace tree multiplier, shown in Figure 5, was designed for this project because it is much faster than the conventional multipliers. The following table summarizes the functions of the various input and output parameters of this multiplier.

Input Parameter	Function Description
A[A0 (LSB) - A7 (MSB)]	8 bits multiplier
B[B0 (LSB) - B7 (MSB)]	8 bits multiplicand

Output Parameter	Function Description
Q [Q0(LSB) - Q15(MSB)]	16 bits product

SIMULATION RESULTS

Simulation results for the Wallace tree multiplier are shown in Figure 6, and found to be functionally correct. For example, for the hexadecimal number system, the simulation results in Figure 6 (a) show that when the multiplier and multiplicand are both 80 H (128 in decimal system), the output is 4000 H (16384 in decimal).

$$\begin{aligned} Q &= A * B; \\ &= 80 \text{ (128 in decimal)} * 80 \text{ (128 in decimal)} \\ &= 40000 \text{ (16384 in decimal)} \end{aligned}$$

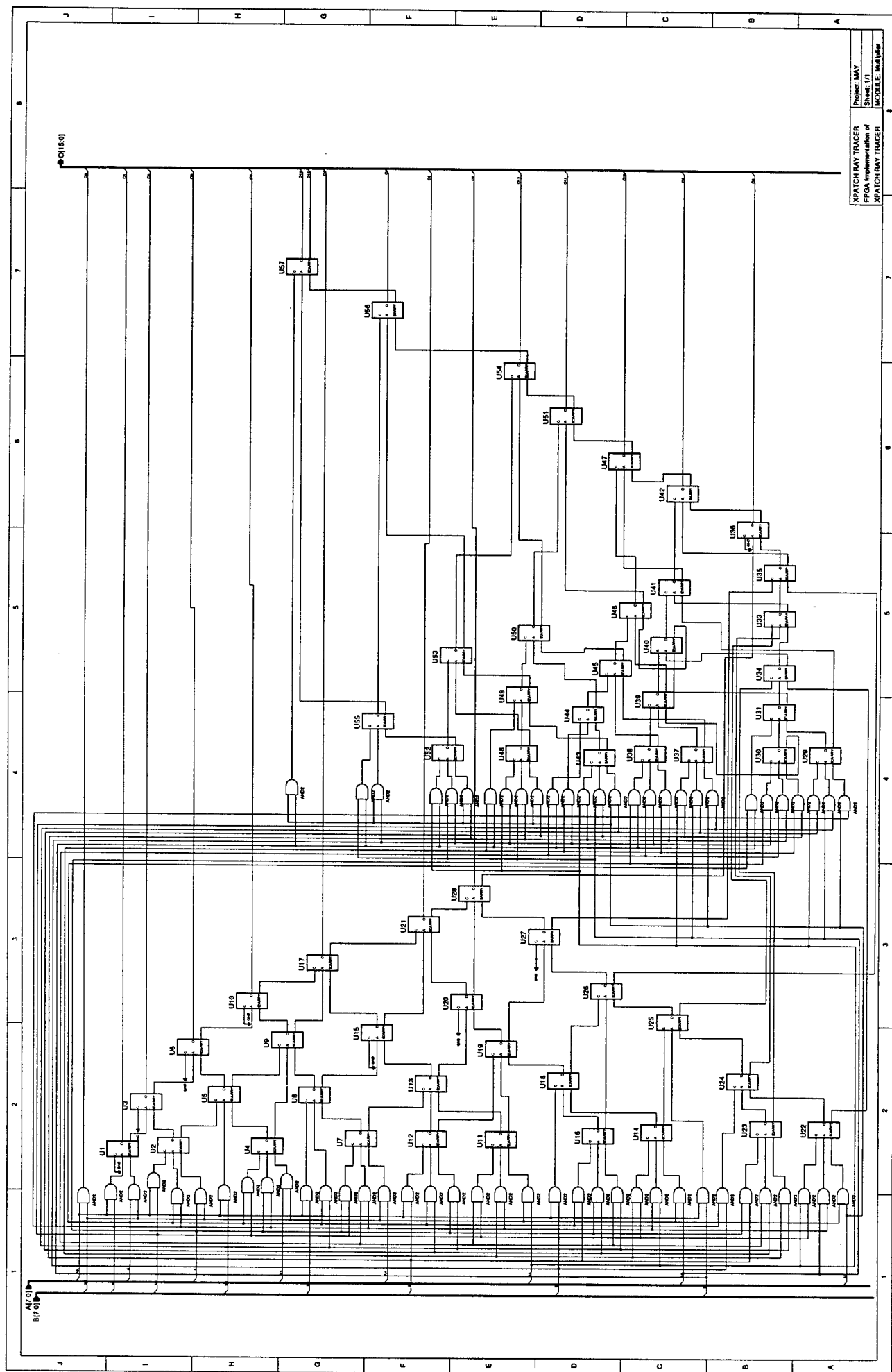


Figure 5: Wallace Tree Multiplier

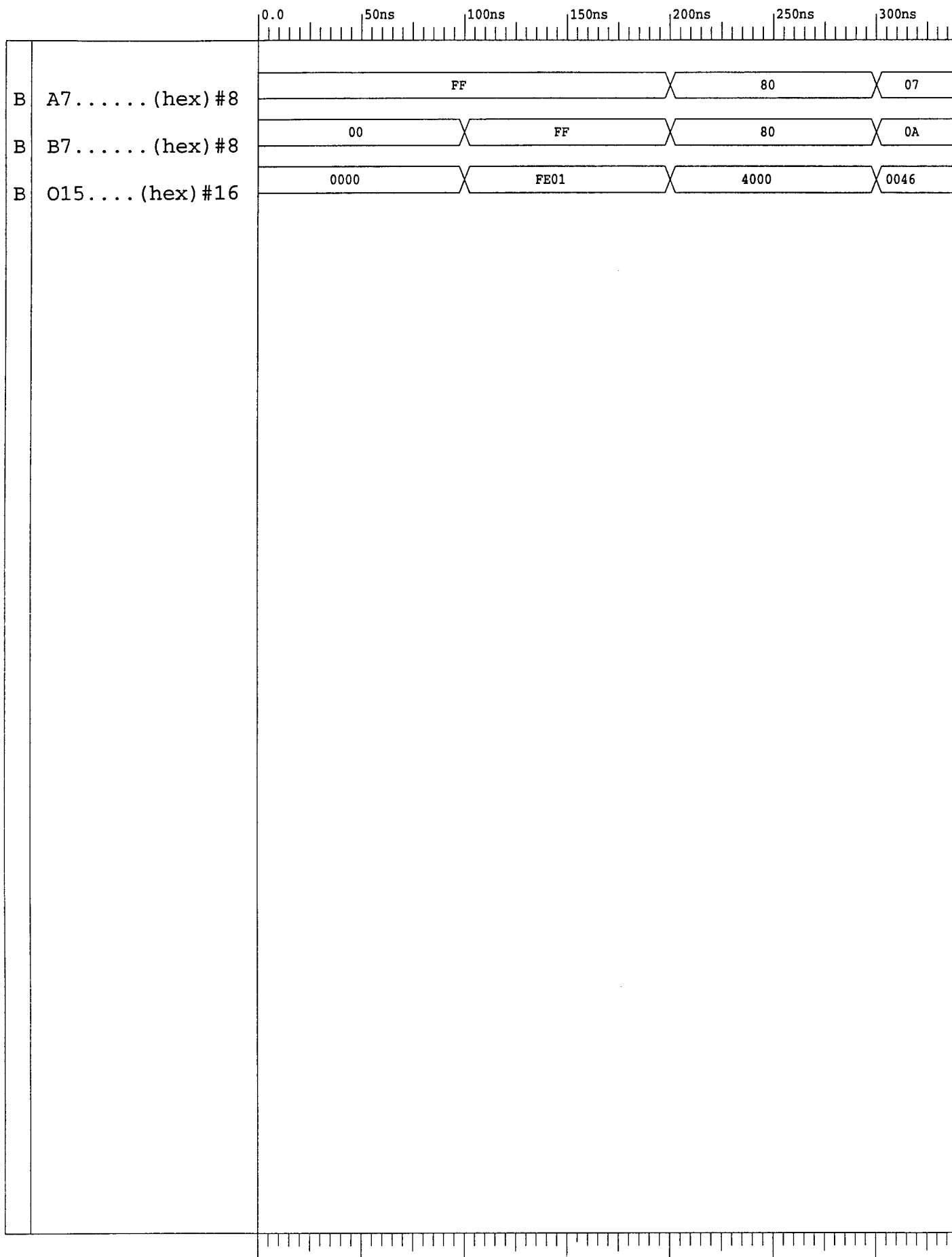
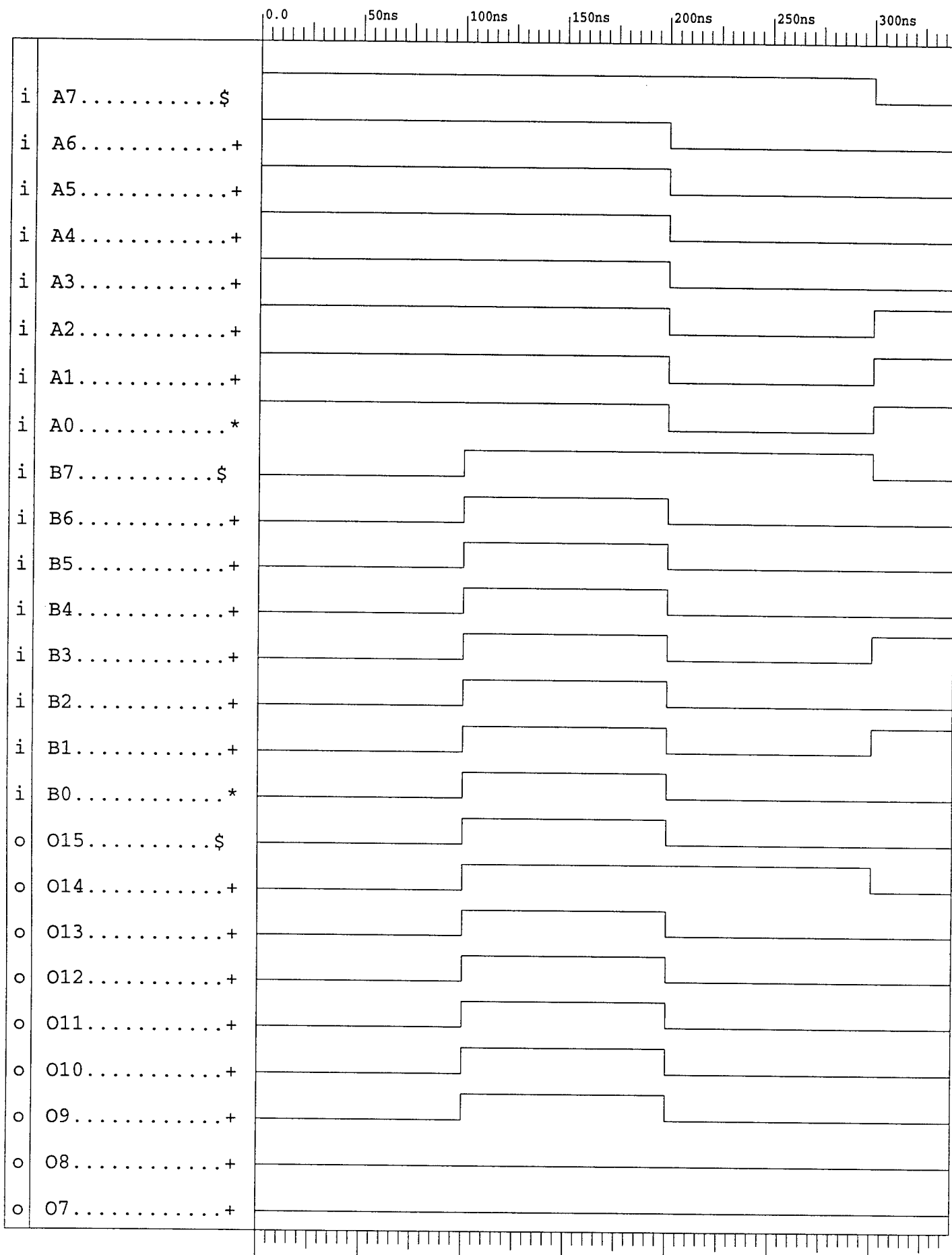
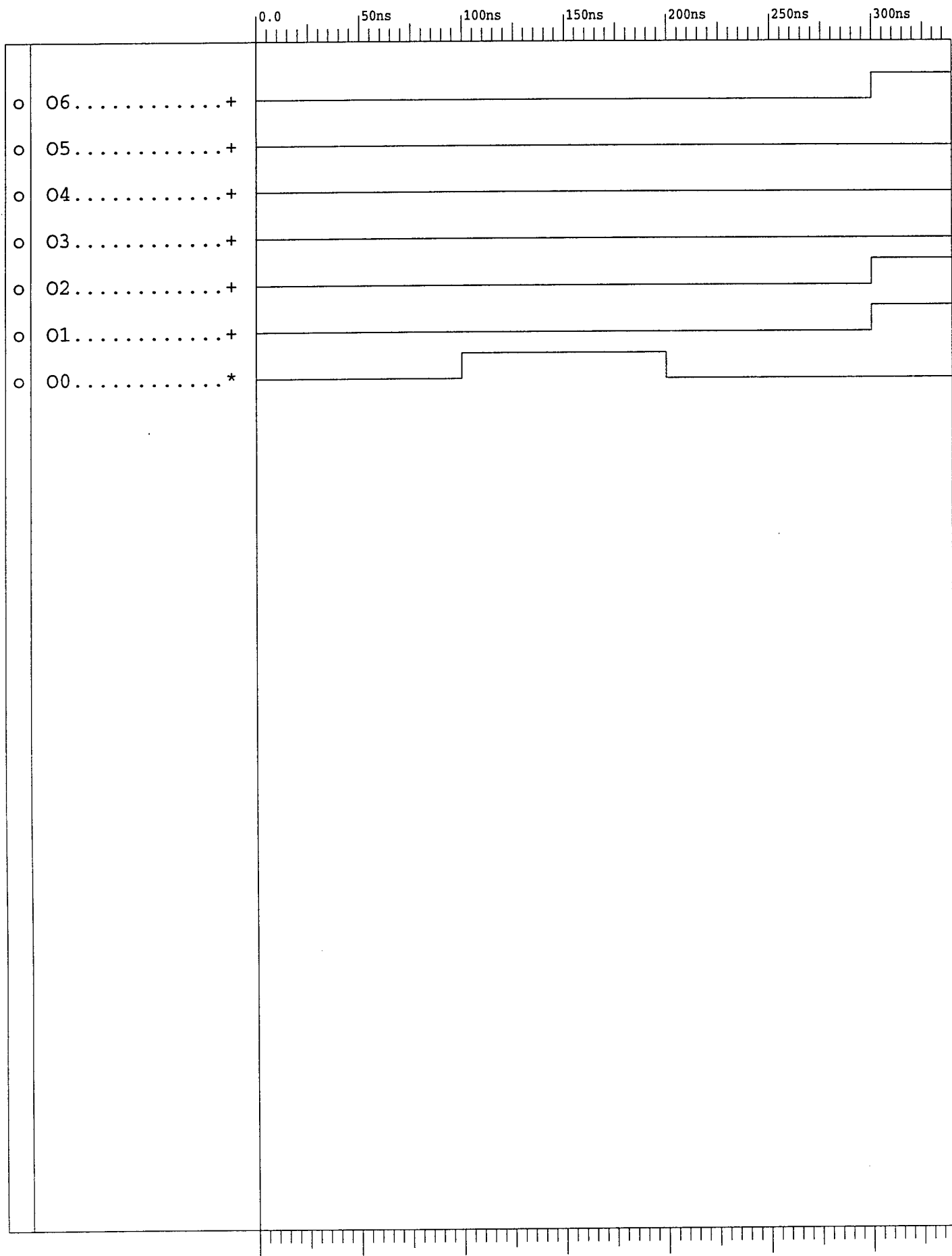


Figure 6(a): Test Simulation (Hex)



20-18 Figure 6(b): Test Simulation (Bin)



3. Module: Parallel in - Parallel out Shift Right Register

The SHIFTR module, shown in Figure 7, depicts our design for an 8 bit parallel in - parallel out shift right register. The components of this module include eight D type flip flops and eight two input multiplexers. The function of this module is to shift one bit of information to the right at each pulse of the clock. The following table summarizes the functions of the various inputs and outputs of this module.

Input Parameter	Function Description
D [D0 (LSB) - D7(MSB)]	8 bits input bus
CLK	Clock function
ENB	Enable function
RESET	Reset function
SHIFT	Shift right function

Output Parameter	Function Description
Q [Q0(LSB) - Q7(MSB)]	8 bits output bus

SIMULATION RESULTS:

The simulation results are shown in Figure 8 and found to be functionally correct. For example, for the input pattern '10000000', the output pattern in Figure 8 (b) is found to be 01000000, 00100000, 00010000, and so on, for successive clock cycles. In the hex system, as shown in Figure 8 (a), the input pattern corresponds to 80 H, and the successive outputs to 40 H, 20 H, and 10 H.

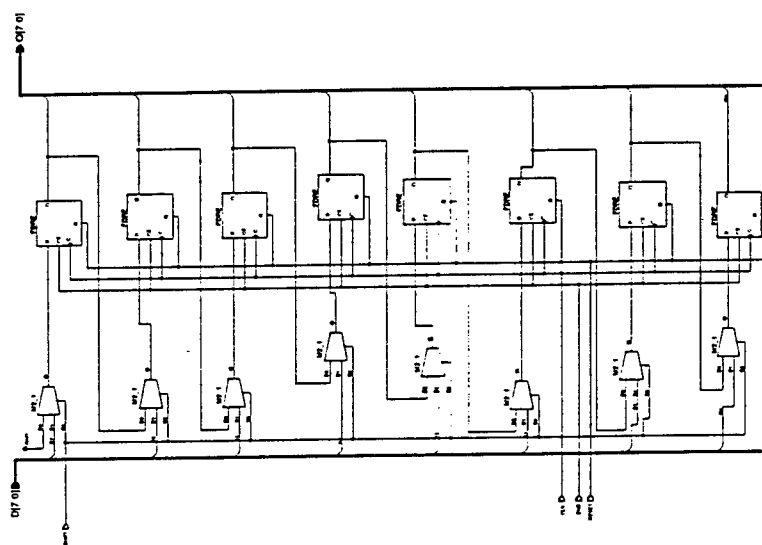
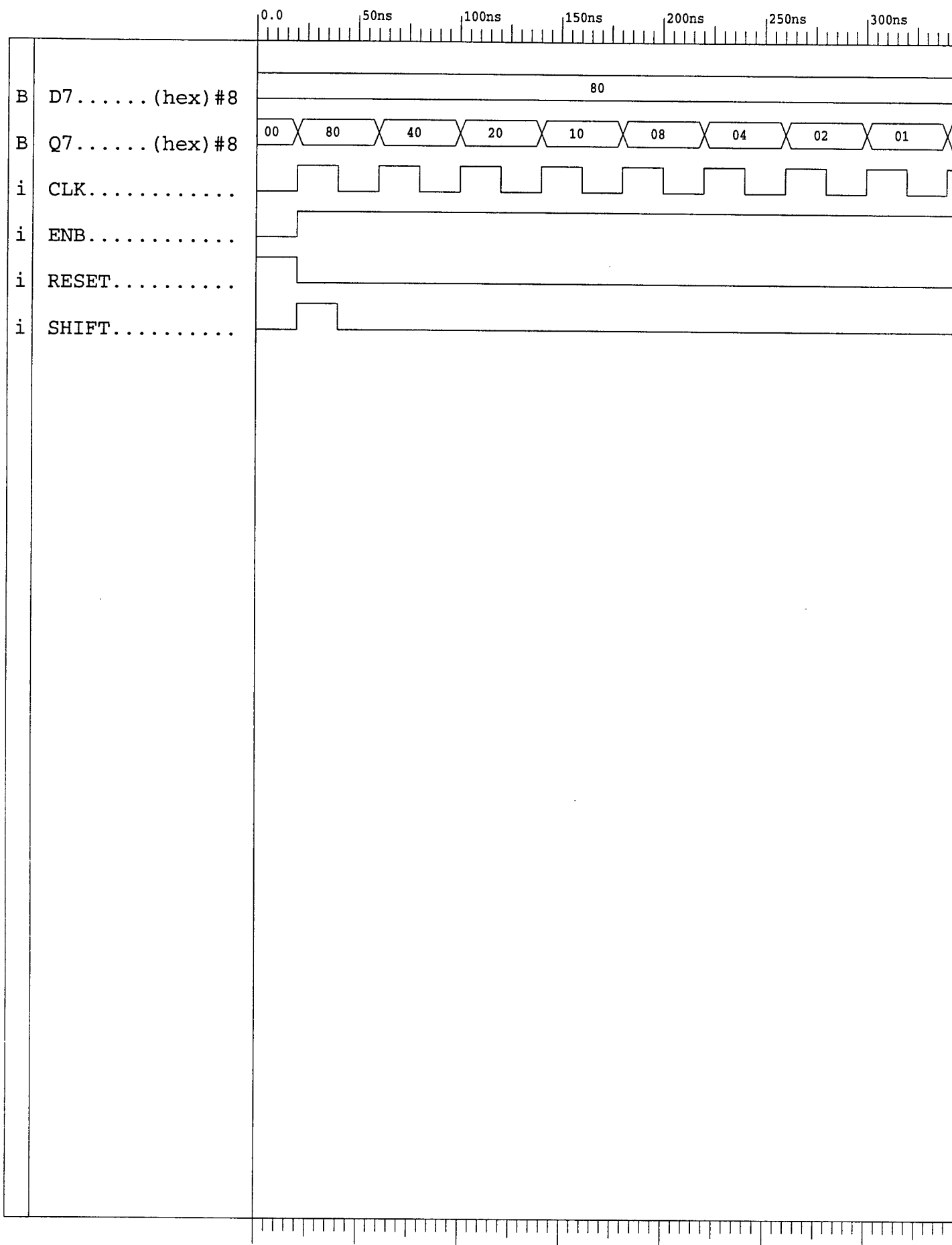
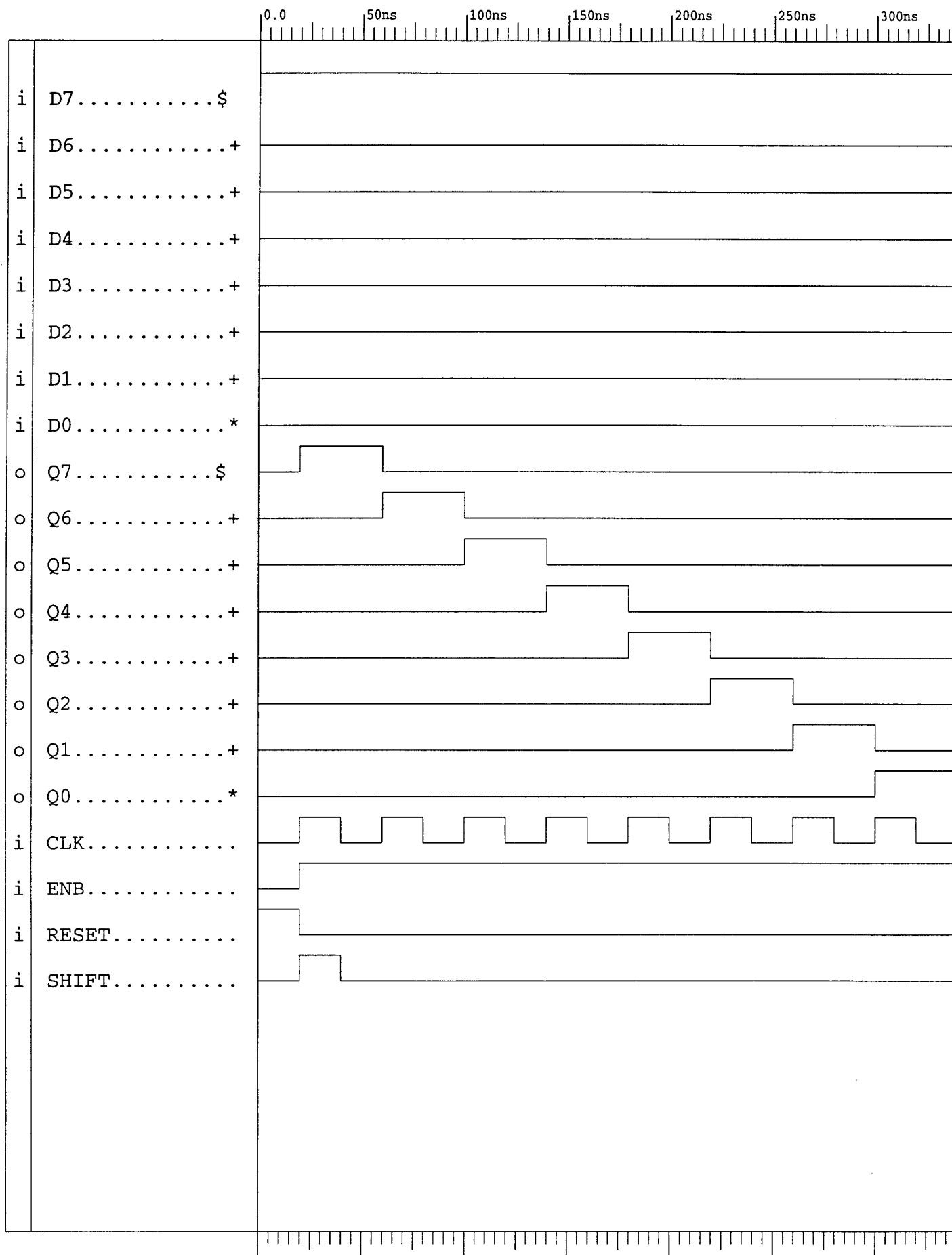


Figure 7: Shift Right Register

WATCHDOG TIMER	POWER MAY
FROM INPUTS OF	BEHOLD
WATCHDOG TIMER	BEHOLD





20-23 Figure 8(b): Test Simulation (Bin)

4. Module: Parallel in - Parallel out Shift Left Register

The SHIFTL module is an 8 bit parallel in - parallel out shift left register. The following table lists the functions of the various input and output parameters of this module.

Input Parameter	Function Description
D [D0 (LSB) - D7(MSB)]	8 bits input bus
CLK	Clock function
ENB	Enable function
RESET	Reset function
SHIFT	Shift left function

Output Parameter	Function Description
Q [Q0(LSB) - Q7(MSB)]	8 bits output bus

SIMULATION RESULTS:

The simulation results shown in Figure 10 verify the functional operation of the shift left register. The simulation results are similar to the previous case except that the data in the input pattern are shifted left instead of right.

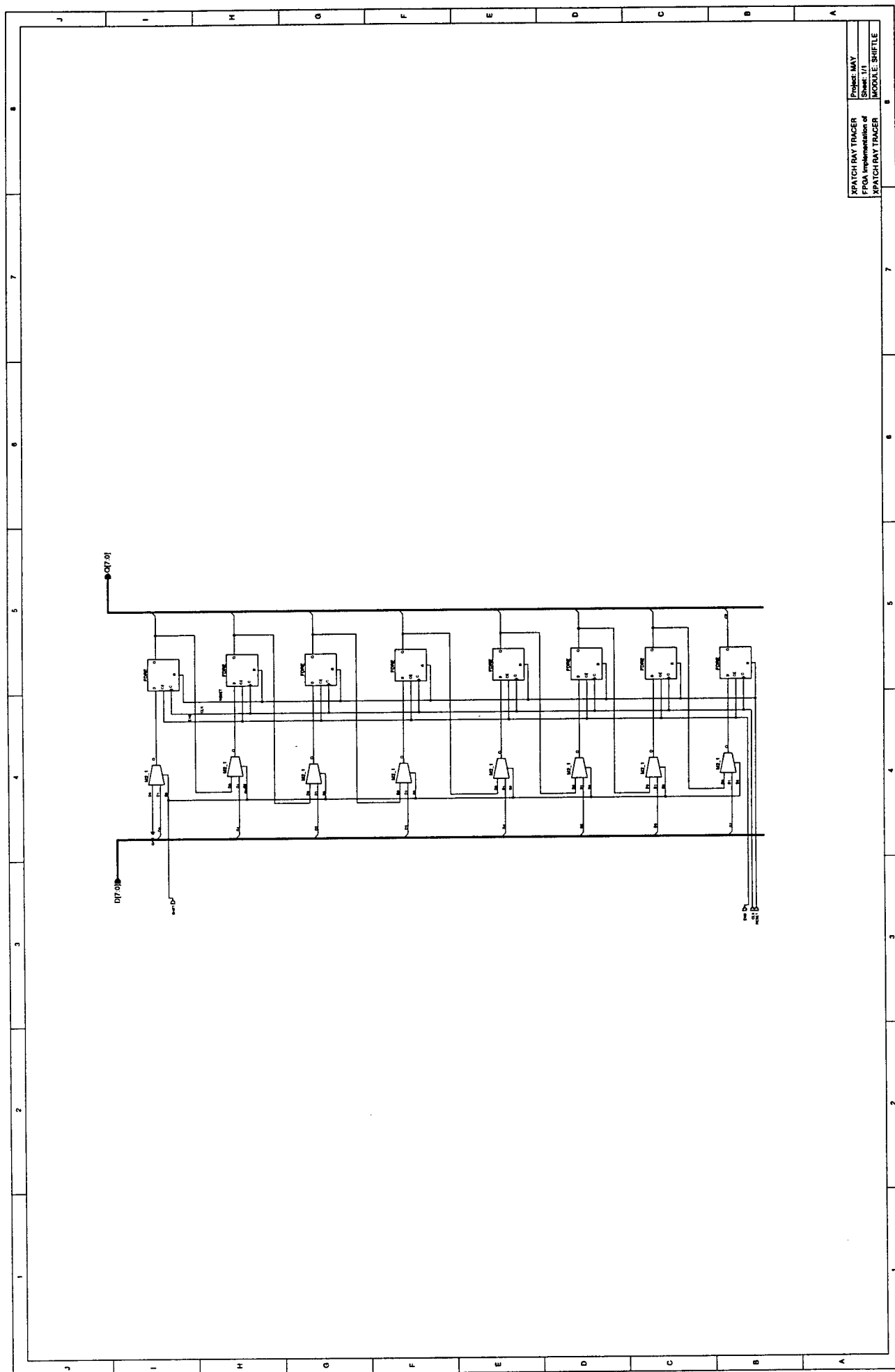
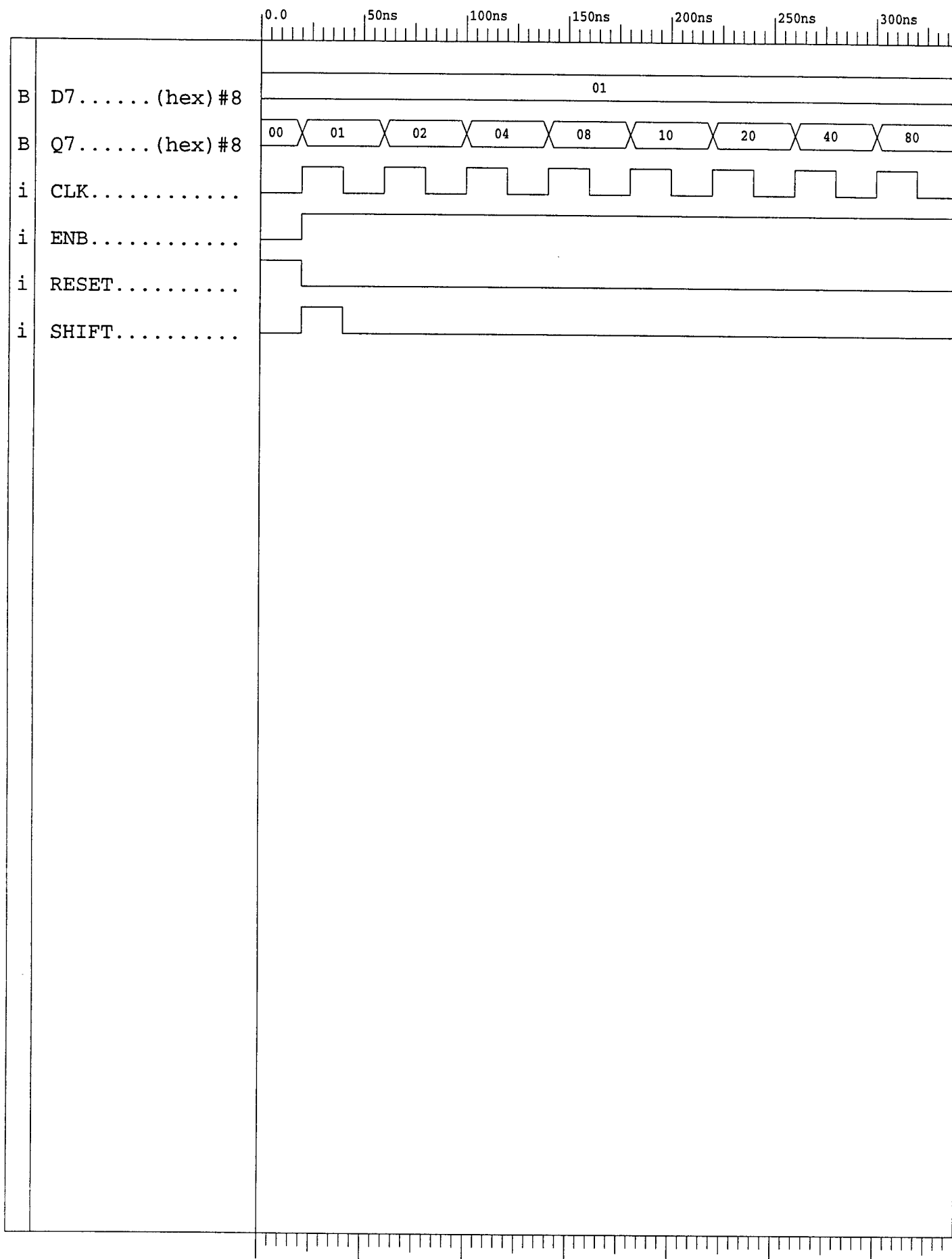
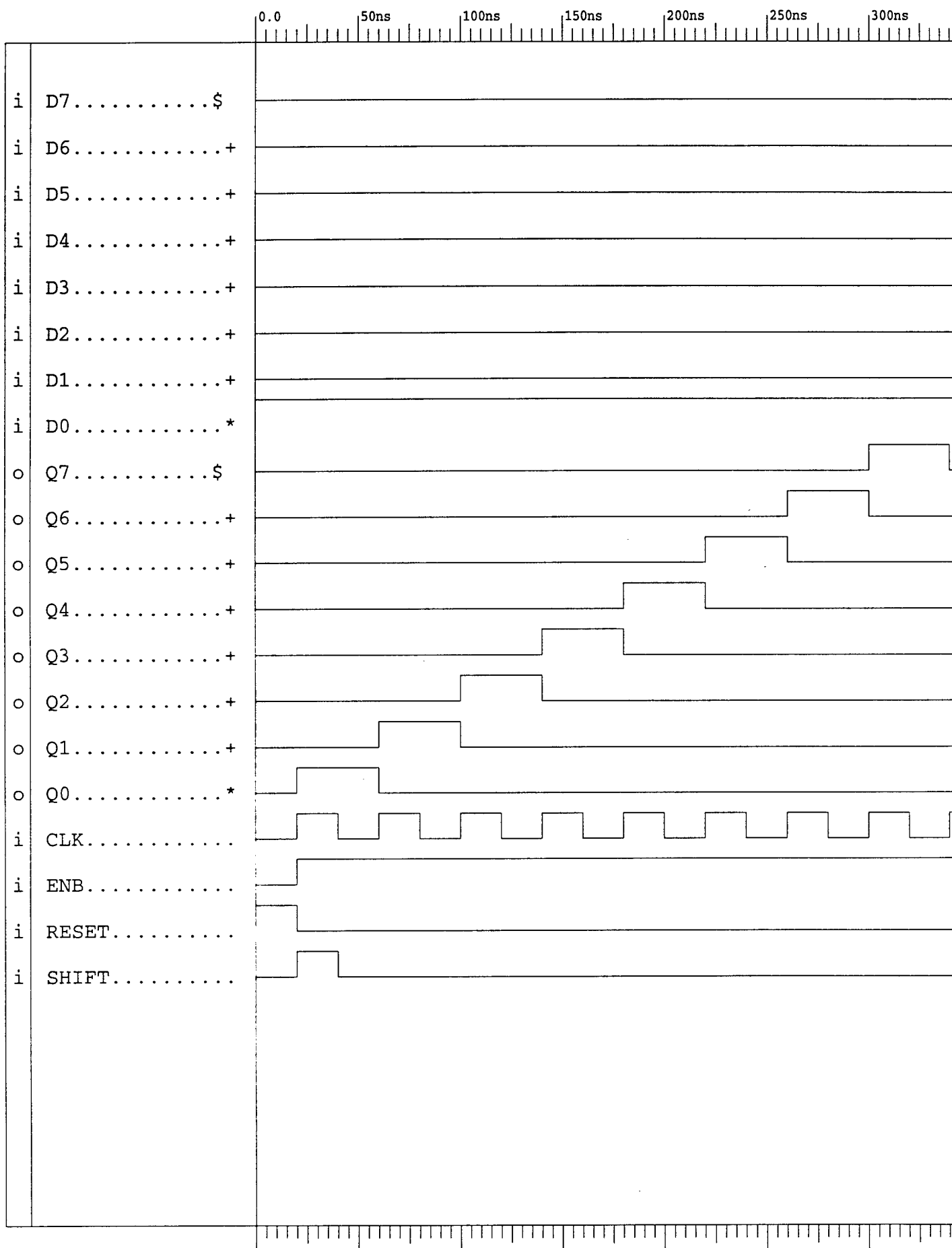


Figure 9: Shift Left Register





20-27 Figure 10(b): Test Simulation (Bin)

5. Module: 2's Complement

The TWCOM8, shown in Figure 11, is an 8-bit 2's complement module. The function of this module is to convert an eight bit input to the 2's complement form. The result is used for the subtraction process. The following table lists the functions of the various inputs and outputs of this module.

Input Parameter	Function Description
A [A0 (LSB) - A7(MSB)]	8 bits input bus

Output Parameter	Function Description
O [0(LSB) - 7(MSB)]	8 bits output bus

SIMULATION RESULTS:

The simulation results for the 2's complement module are shown in Figure 12, and found to be functionally correct. For example, it is seen from Figure 12 that at 150 ns, the input is 71 H and the corresponding output 8F, as expected.

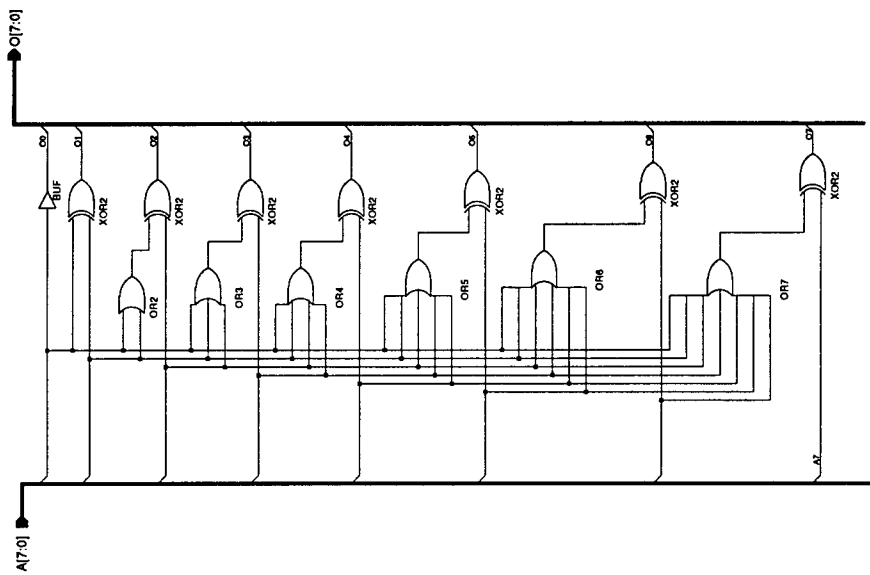


Figure 11: 2's Complement

XPATCH RAY TRACER	Project: MAY
FPGA Implementation of	Sheet: 1/1
XPATCH RAY TRACER	MODULE: 2's Complement

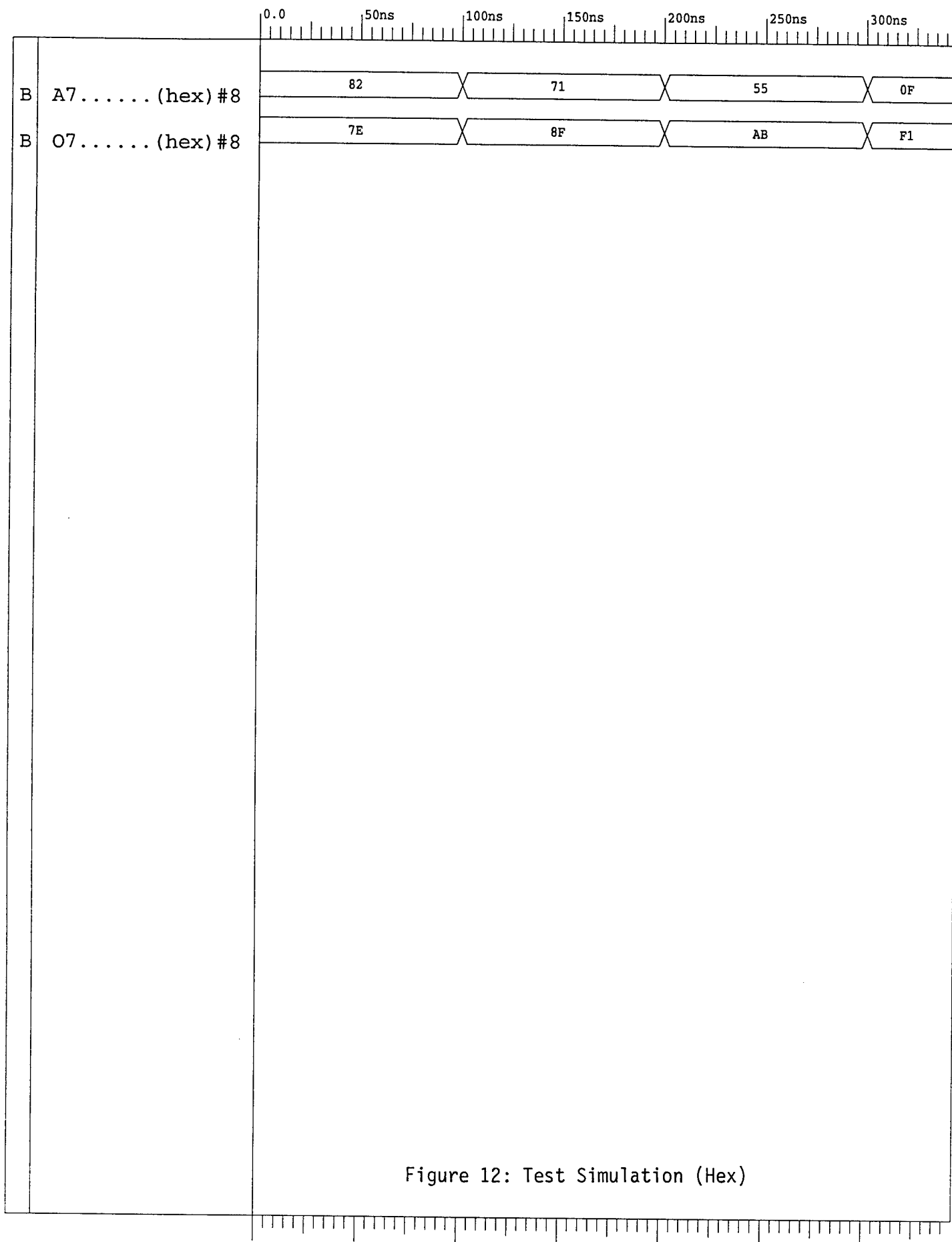


Figure 12: Test Simulation (Hex)

The Systolic Processing Element:

The schematic for the systolic processing element is shown in Figure 13. The complete schematic of the systolic array, corresponding to the architecture proposed earlier in Figure 2, is shown in Figure 14. As can be seen from this figure, the systolic array comprises of nine identical systolic processing units.

The systolic processing element comprises of the various modules described in the previous sections. The function of the systolic processing element is to compute the 't' parameter. The 't' parameter, as explained earlier, is given by:

$$t = [BBP - rs] / rd$$

In the schematic for the systolic processing element shown in Figure 13, input signal 'A' corresponds to 'BBP', and input 'B' corresponds to the parameter 'rs'. The inverse of 'rd' is represented by 'M', a floating point number. Signal 'Ein' represents the value of the exponent for the floating point number. The following table lists the functions of the various input and output parameters of the systolic processing element.

Input Parameter	Function Description
A [A0 (LSB) - A7(MSB)]	8 bits input bus
B [B0(LSB) - B7(MSB)]	8 bits input bus
EIN1 [EIN0(LSB) - EIN1(MSB)]	Exponent inputs
CLK	Clock function
ENB	Enable function
RESET	Reset function
SHIFT	Shift left function

Output Parameter	Function Description
O [O0(LSB) - O15(MSB)]	16 bits output bus
EOUT [EOUT0(LSB) - EOUT1(MSB)]	Exponent output

SIMULATION RESULTS:

The simulation results for the systolic processing element is shown in Figure 15, and found to be functionally correct. For example, for the time range 225ns to 300ns, the inputs are: A = 29 H (41 in decimal), B = 0E H(15 in decimal), M = 09 (9 in decimal), and Ein = 1. In other words, we need to compute: $(41 - 15) * 0.9 = 23.4$. By checking the simulation results, we verify the functional operation of the processing element as follows:

sum = A-B = 29 H (41 in decimal) - 0E H(15 in decimal) = 1B(26 in decimal).

Ein = 1 (which means that M should be multiplied by 0.1)

M = 9

O15 = sum * M = 1B H(26 in decimal) * 09 H (9 in decimal)

= 00F3 H(234 in decimal)

Eout = 1 (which means that O15 equals to 23.4 in decimal, as expected)

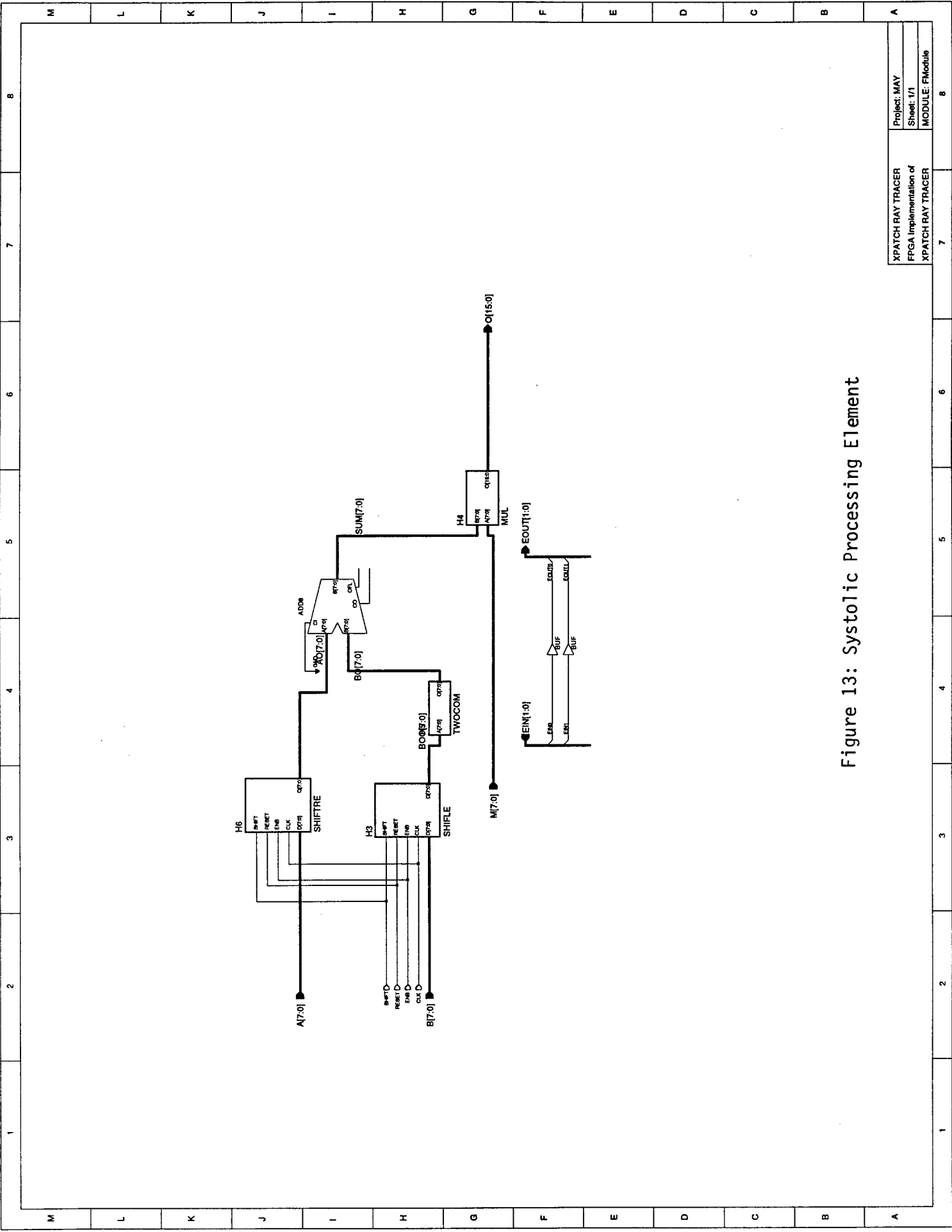
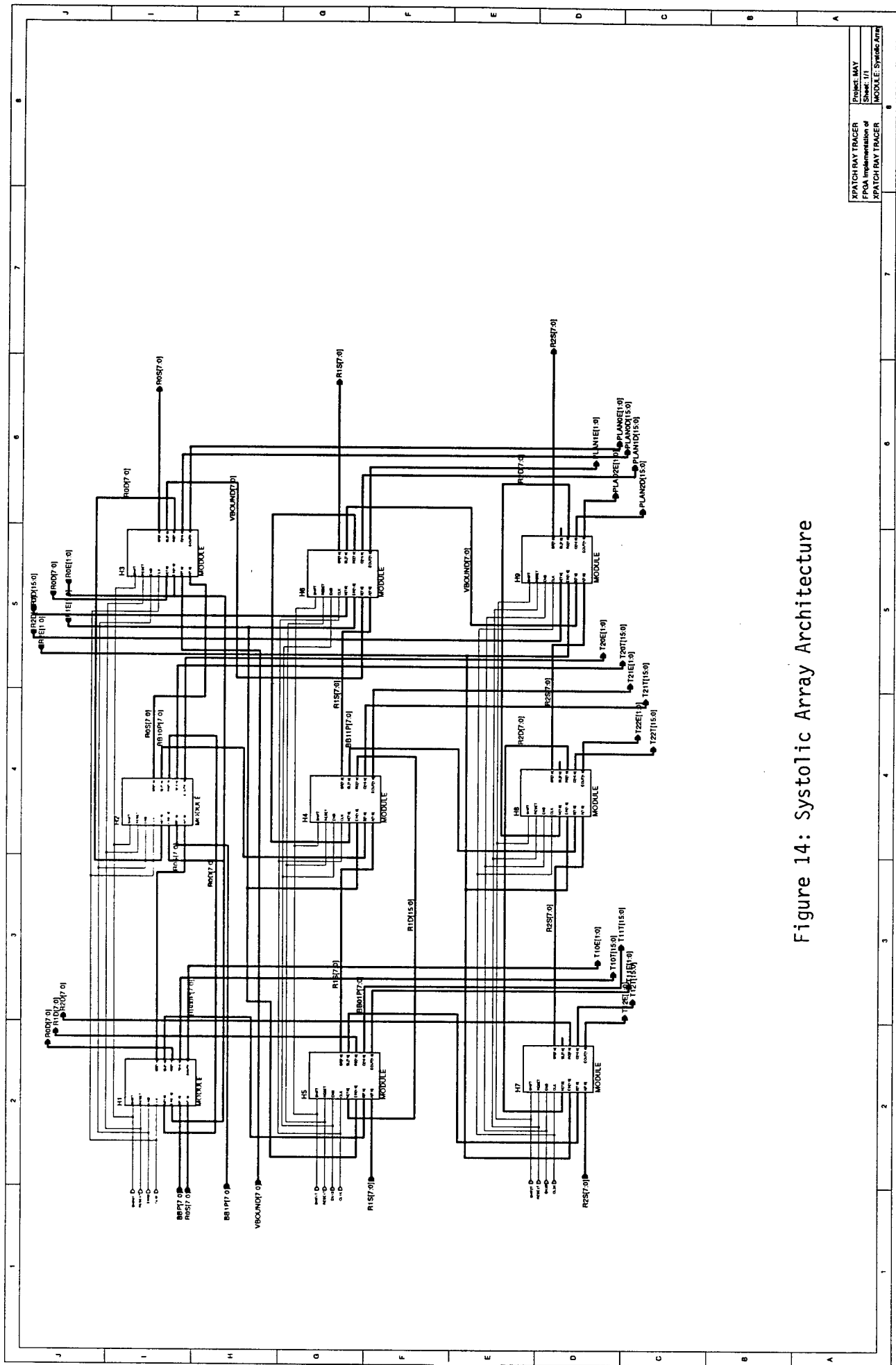


Figure 13: Systolic Processing Element

XPATCH RAY TRACER	Project: MAY
FPGA Implementation of	Sheet: 1/1
XPATCH RAY TRACER	MODULE: FModule



XPATCH RAY TRACER	Project: MAY
FPGA Implementation of	Sheet: 1/1
XPATCH RAY TRACER	Module: Systolic Array

Figure 14: Systolic Array Architecture

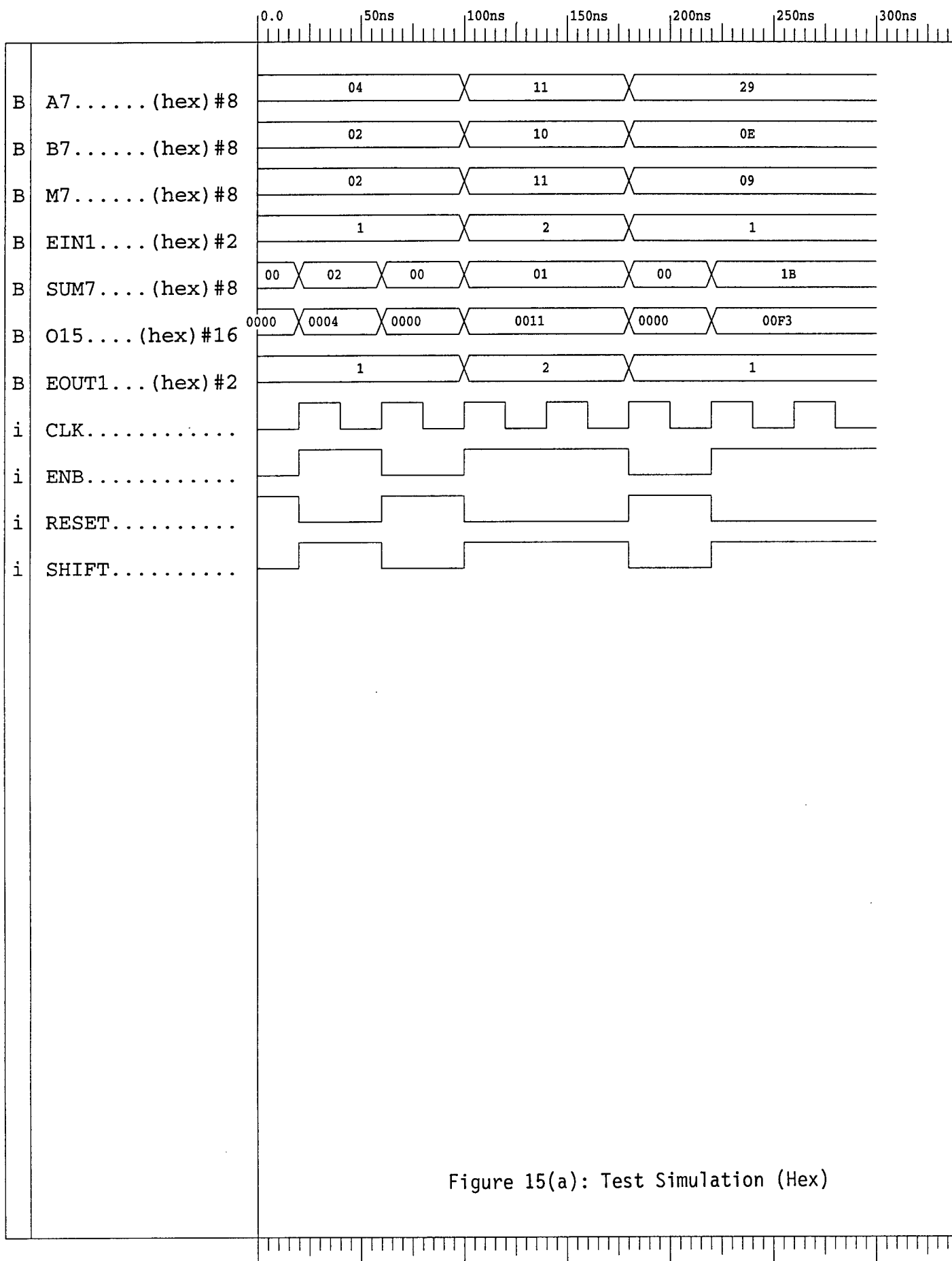
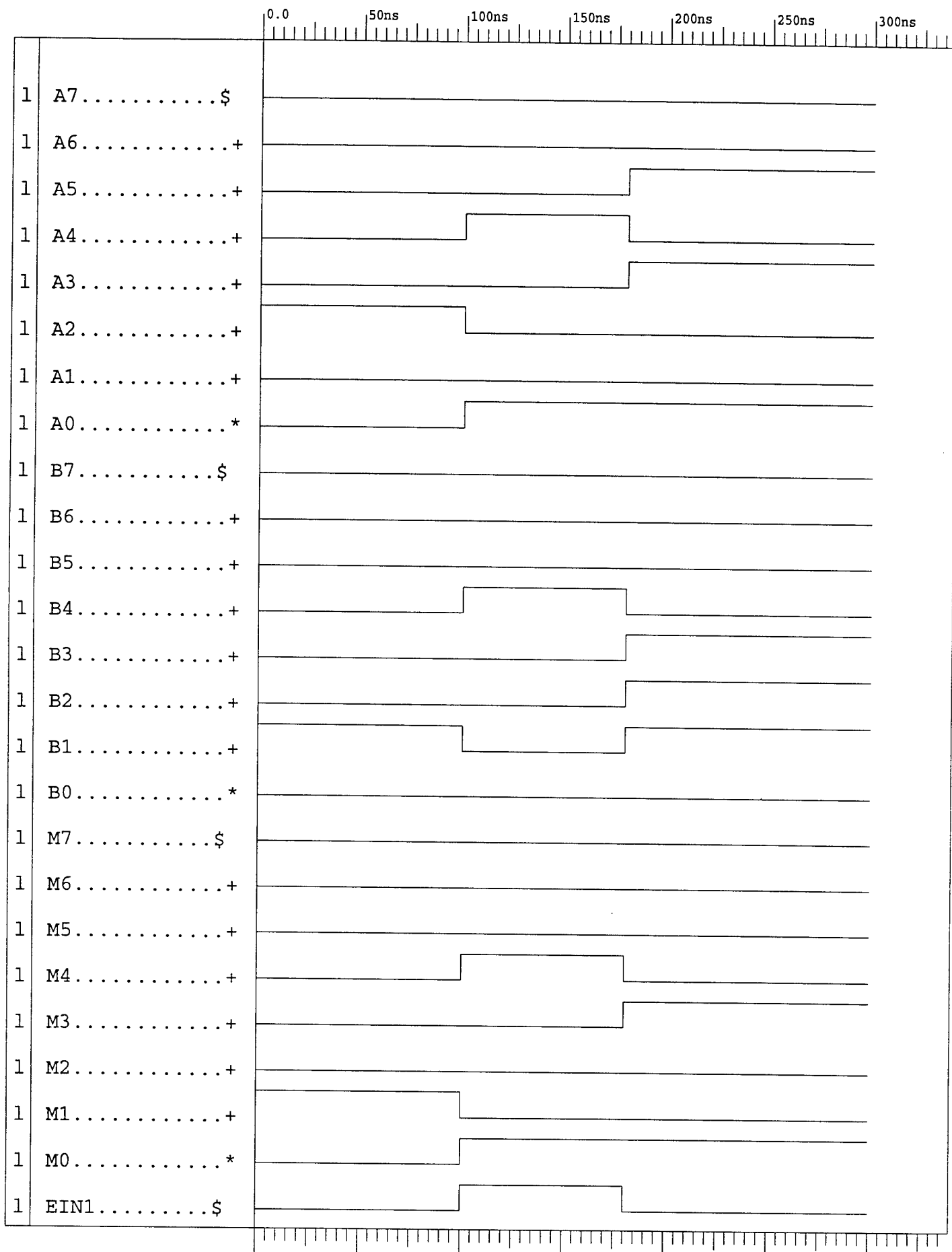
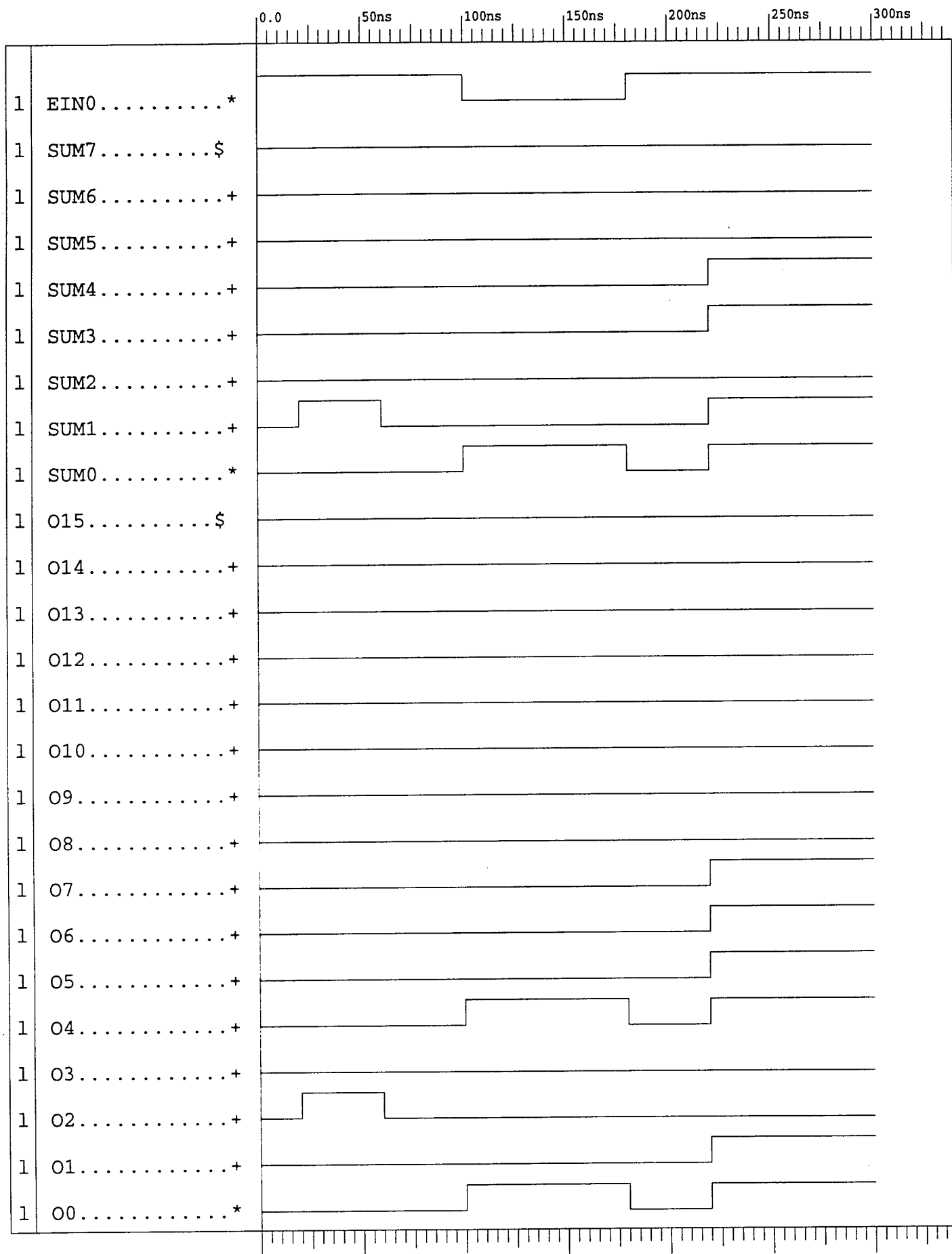
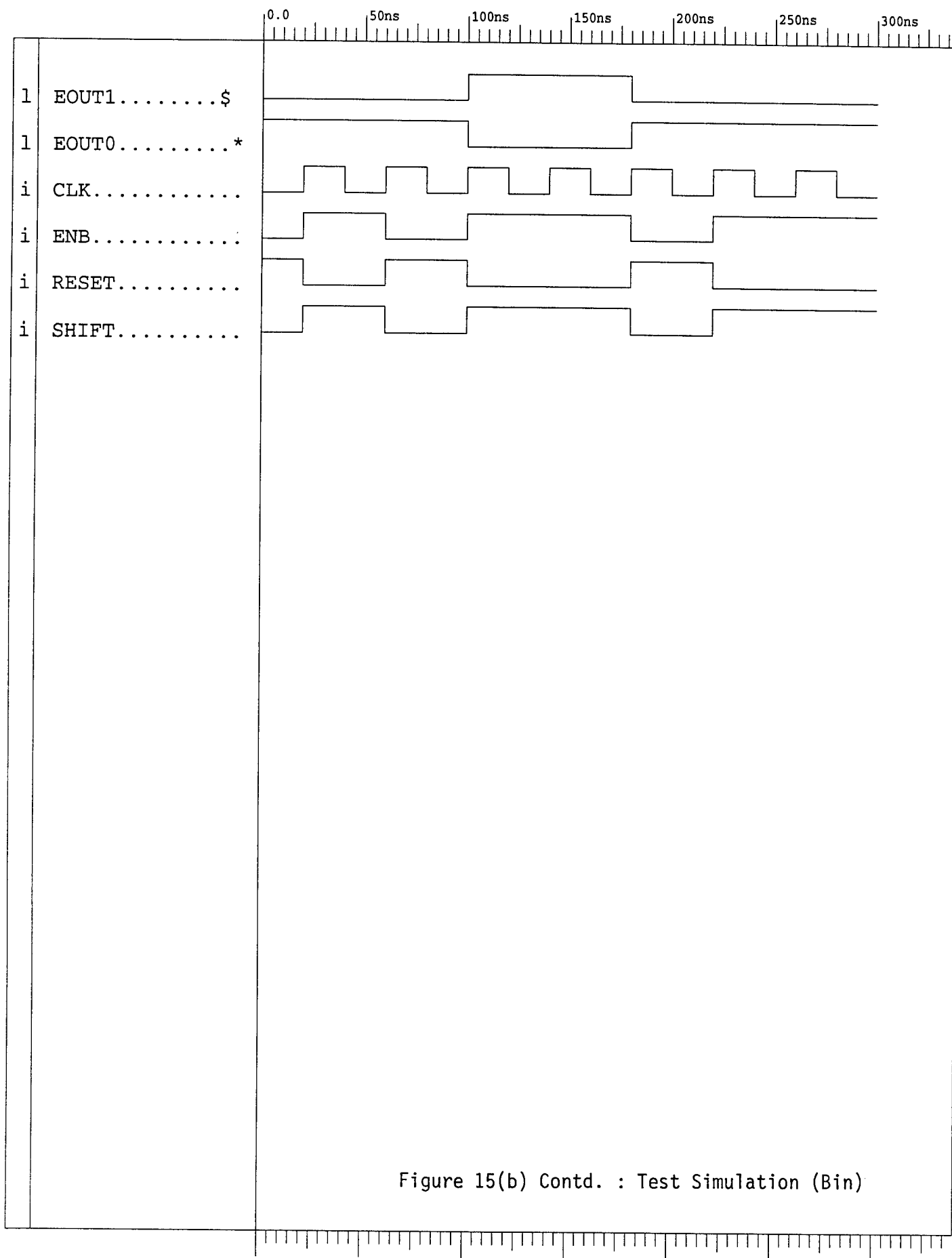


Figure 15(a): Test Simulation (Hex)





20-37 Figure 15(b) Contd. : Test Simulation (Bin)



Implementation Results

The implementation results, in terms of CLB (configurable logic blocks) and IOB (input output buffers) utilization are shown in the following tables. These results have been generated using the Xilinx M1 Foundation tools. It is seen from the table, that the systolic processing module occupies 10 % of the xc4025E FPGA. Since there are nine identical processing elements in our systolic array architecture, the entire systolic array will roughly occupy 90% of the CLBs in the xc4025E FPGA. In terms of IOBs, it is seen that the systolic module requires 18 % of the IOBs on the xc4025E. A multiplication factor of 9 (for the nine systolic processing elements) for the entire systolic array yields an utilization factor of 162 %. Therefore, two daisy chained xc4025E FPGAs are required for mapping the complete systolic array architecture.

CLB Utilization for select Xilinx FPGAs

	xc4003E	xc4005E	xc4006E	Xc4008E	xc4010E	xc4013E	xc4020E	xc4025E
Add8 Alone	6/100 (6%)	6/196 (3%)	6/256 (2%)	6/324 (2%)	6/400 (2%)	6/576 (1%)	6/784 (1%)	6/1024 (0%)
ShiftLe Alone	8/100 (8%)	8/196 (4%)	8/256 (3%)	8/324 (2%)	8/400 (2%)	8/576 (1%)	8/784 (1%)	8/1024 (0%)
ShiftRi Alone	8/100 (8%)	8/196 (4%)	8/256 (3%)	8/324 (2%)	8/400 (2%)	8/576 (1%)	8/784 (1%)	8/1024 (0%)
TwoCom Alone	5/100 (5%)	5/196 (2%)	5/256 (1%)	5/324 (1%)	5/400 (1%)	5/576 (0%)	5/784 (0%)	5/1024 (0%)
Mul Alone	85/100 (85%)	85/196 (43%)	85/256 (33%)	85/324 (26%)	85/400 (21%)	85/576 (14%)	85/784 (10%)	85/1024 (8%)
Systolic Module	Cannot Fit	112/196 (57%)	112/256 (43%)	112/324 (34%)	112/400 (28%)	112/576 (19%)	112/784 (14%)	112/1024 (10%)

IOB Utilization for select Xilinx FPGAs

	xc4003E	xc4005E	xc4006E	Xc4008E	xc4010E	xc4013E	xc4020E	xc4025E
Add8 Alone	27/80 (33%)	27/112 (24%)	27/128 (21%)	27/144 (19%)	27/160 (17%)	27/192 (14%)	27/224 (12%)	27/256 (11%)
ShiftLe Alone	19/80 (23%)	19/112 (16%)	19/128 (14%)	19/144 (13%)	19/160 (11%)	19/192 (9%)	19/224 (8%)	19/256 (7%)
ShiftRi Alone	19/80 (23%)	19/112 (16%)	19/128 (14%)	19/144 (13%)	19/160 (11%)	19/192 (9%)	19/224 (8%)	19/256 (7%)
TwoCom Alone	16/80 (20%)	16/112 (14%)	16/128 (12%)	16/144 (11%)	16/160 (10%)	16/192 (8%)	16/224 (7%)	16/256 (6%)
Mul Alone	32/80 (40%)	32/112 (28%)	32/128 (25%)	32/144 (22%)	32/160 (20%)	32/192 (16%)	32/224 (14%)	32/256 (12%)
Systolic Module	Cannot Fit	47/112 (41%)	47/128 (36%)	47/144 (32%)	47/160 (29%)	47/192 (24%)	47/224 (20%)	47/256 (18%)

Acknowledgments

The author wishes to acknowledge the financial support given by AFOSR for making this work possible. At the Wright Laboratory Avionics Directorate, several persons assisted the author in developing this project. Foremost, the author wishes to thank Kerry Hill for defining and supervising the entire project. Without her help, this project would not have been possible. Thanks are also due to Brian Kadrovach, Jeff Summers, and Tim Read for assisting in this project. Thanks are also due to Darrell Barker for serving as the team leader in the various group meetings held to monitor the progress of the project during the summer of 1996.

At the University of Toledo, graduate student Hung Lin assisted in the successful completion of this project. Her assistance is gratefully acknowledged.

References

1. Andrew Glassner, An Introduction to Ray Tracing, Academic Press, 1989.
2. H. Ling, R. Chou, and S.W. Lee, "Shooting and Bouncing Rays: Calculating the RCS of an Arbitrarily Shaped Cavity," IEE Trans. On Antennas and Propagation, vol. 37, 194 - 205, 1988.
3. Users Manual for XPATCH, Demaco, 100 Trade Center Drive, Champaign, IL, 1993.
4. E.M. Miller, D.J. Andersh, and A.J. Terzzouli Jr., "Facetization Level and the Effect on XPATCH Predictions," Proceedings of the Ninth Annual Review of Progress in Applied Computational Electromagnetics, Naval Postgraduate School, Monterey, CA, March 22 -26, 1993, pp. 610 - 617.
5. J.T. Kajiya, "Siggraph 83 Tutorial on Ray Tracing," Siggraph 83 State of the Art in Image Synthesis Course Notes, July 1983.

**DEVELOPMENT OF OPTIMUM DRILLING PROCESS FOR
ADVANCED COMPOSITES**

Anthony C. Okafor
Associate Professor
Department of Mechanical and Aerospace Engineering
and Engineering Mechanics

University of Missouri-Rolla
Rolla, Missouri, 65409-0050

Final Report for:
Summer Research Extension Program
and
Wright Laboratory
Wright Patterson Air Force Base
Dayton, Ohio 45433

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and
Wright Laboratory

June 30, 1998

DEVELOPMENT OF OPTIMUM DRILLING PROCESS FOR ADVANCED COMPOSITES

Anthony C. Okafor
Associate Professor
Department of Mechanical and Aerospace Engineering
and Engineering Mechanics
University of Missouri-Rolla

Abstract

Extensive experimental investigation was conducted to study the effect of drilling conditions (speed and feedrate), drill material and drill point angle on thrust force, torque, acoustic emission, drilled hole quality (hole size, roundness, and delamination), and drill wear during drilling of advanced composites. The overall objective of the research is to develop optimum drilling process and an intelligent monitoring system for predicting drill wear and drilled hole quality in advanced composites. A 1/3 fractional factorial design of experiment with 4 factors at 3 levels was used to conduct the series of drilling experiments. Acoustic emission, cutting force, and cutting torque were monitored during drilling in advanced composite plates. The acquired sensor data along with measured drill wear were used to determine optimum drilling condition to develop smart sensing system using neural network for assuring the quality of drilled holes in advanced composites. The project is divided into four major tasks. The first task involved conducting series of drilling experiments using the 1/3 fractional factorial design of experiment, monitoring, acquiring, and processing the thrust force, drilling torque, and acoustic emission data, hole exit delamination, hole diameter and hole roundness produced during drilling in advanced composite plate (AS4/PEEK). The second task involved characterizing the main effects and two factor interaction of cutting conditions (feedrate and speed), drill point angle, and drill tool materials on the drilling process response (thrust force, drilling torque, acoustic emission, hole exit delamination, hole diameter, and hole roundness), and identification of optimum drill tool material, drill point angle, and cutting conditions for delamination free hole drilling in advanced composites. Three drill materials - High Speed Steel Cobalt, Carbide, and Polycrystalline Diamond (PCD) were investigated. For the third task, additional drilling experiments were conducted using the identified optimum drilling condition to study drill wear and drill life characteristics for two of the identified two best drill materials (carbide and PCD). SEM photographs of drilled specimen showing extent of delamination and internal cracks were taken and used to corroborate the analysis from sensor data. The fourth task involved the development of neural network models for predicting delamination, drill wear, hole diameter and roundness based on the integration of the multi-sensor data. This fourth task is the subject for future research. The composite plate material investigated is a carbon fiber reinforced composites (AS4/PEEK) supplied by McDonnell Douglas Aircraft Company in St. Louis, Missouri. The experiment were performed on a Cincinnati Milacron Sabre 750 Machining Center.

TECHNICAL INFORMATION

A. Summary of Results

One MS thesis on drilling of advanced composites and three papers [1-3] were obtained from the research and are in the final preparation stages for publication. This research investigates the effect of cutting condition, drill material, and drill point angle in drilling of advanced composites. HSS Cobalt, Solid Carbide, and Polycrystalline Diamond (PCD) Drills were investigated. In reference [1-2], a 1/3 Fractional Factorial Design of Experiment with 4 factors at 3 levels was used to determine the main effects, as well as two factor interaction of drill point angle, cutting speed, feed rate, and drill material on drilling thrust force, drilling torque acoustic emission, and drill hole quality (hole exit delamination, hole diameter, and hole roundness). In reference [3], the optimal drilling conditions were selected to maximize hole quality and additional holes were drilled to study the effects of tool wear on thrust force, torque, and acoustic emission. The tool wear portion of the experiment was performed using both a solid carbide drill and a polycrystalline Diamond (PCD) Drill. Thus the research resulted to three major papers which are now under review for journal publications. These three papers are summarized in section E in the form of abstract and conclusions.

B. Scientific Collaborators:

Dr. A. C. Okafor, Associate Professor and Principal Investigator

Dr. Wen F. Lu, Associate Professor and Co-Principal Investigator

Mr. Scott Birdsong, an MS student and research assistant who will be completing his degree at the University of Missouri-Rolla in August, 1998.

Mr. Yalcin Ertekin, a Ph.D. student and research assistant who will be completing his degree at the University of Missouri-Rolla in December, 1998.

Mr. Anietie Ukpog, a Ph.D. student and research assistant who will be completing his degree at the university of Missouri-Rolla in December, 1998.

Mr. Yang Yuping, a Ph.D. student and research assistant who completed his degree at the University of Missouri-Rolla in December 1996.

Mr. Emmanuel Ugo Enemuoh, a Ph.D. student and research assistant who will be completing his degree at the University of Missouri-Columbia.

C. Publications:

1. A. Chukwujekwu Okafor and Scott Birsong, "Effect of Drilling Conditions, Drill Material and Point Angle on Thrust Force and Torque in Drilling Advanced Composite Material," under preparation for submission to the Journal of Manufacturing Systems: Special Issue on Manufacturing Processes.
2. A. Chukwujekwu Okafor and Scott Birsong, "Effect of Drilling Conditions, Drill Material and Point Angle on Hole Quality and Delamination in Drilling Advanced Composite Material," under preparation for submission to the Journal of Manufacturing Systems: Special Issue on Manufacturing Processes.
3. A. Chukwujekwu Okafor and Scott Birsong, "Effect of Tool Wear on Thrust Force, Moment and Acoustic Emission in Drilling Advanced Composite Material," under preparation for submission to the Journal of Machine Tools and Manufacture.

D. Theses Completed Acknowledging the Award:

1. Scott Birdsong, MSME, University of Missouri-Rolla, August, 1998, " Machining of Advanced Composites", completed.

E. Title and Abstracts of Three Papers Resulting from the Research:

**I. EFFECTS OF DRILLING CONDITIONS, DRILL MATERIAL, AND POINT ANGLE
ON THRUST FORCE AND TORQUE IN DRILLING ADVANCED COMPOSITE
MATERIAL**

Abstract

A 1/3 fractional factorial design of experiment with 4 factors at 3 levels was used to study drilling of advanced multi-directional carbon fiber reinforced composite plate. Series of 1/4 inch holes were drilled in a 0.3 inch thick AS4/PEEK composite plate as per the 1/3 fractional factorial design of experiment. The plate was mounted on a specially designed fixture which allowed for the acquisition of thrust force, drilling torque (moment) from a Kistler 9272 four component dynamometer. The voltage proportional signals were separated via Kistler Type 5405A junction box and fed into separate Kistler Type 5010 dual mode amplifiers. The amplified signals were ran through low pass filters to reduce the unwanted noise from the drilling process. The digitized signals were save and transferred to a Pentium PC for further processing. Acoustic emission signals generated by the drilling process was picked up using Digital Wave B-1025 broadband sensors mounted on the AS4/PEEK composite plate and to the drilling fixture. The AE signals were passed through Digital Wave PA2040GA AE pre-

amplifiers with 40 dB gain. The voltage proportional analog signals were fed into a Digital wave F4000 Fracture Wave Detector where the signals were filtered with band-pass filters, digitized at a sampling frequency of 6.25 MHz with record length of 2048 data points and saved directly to a Pentium PC. The Thrust force, drilling torque, and acoustic emission signals were acquired for the entire length of each hole drilling cycle. The design of experiment was conducted to determine the main effects and two factor interactions of: drill point angle, cutting speed, feed rate, and drill material on thrust force and torque (drilling moment) in drilling AS4/PEEK composite plate. Drill material, feed rate, and cutting speed were found to have statistically significant effects on thrust force and torque. Drill material, feed rate and cutting speed account for over 78% of the variation measured in the maximum thrust force and RMS thrust force. Cutting speed and drill material were found to be the only statistically significant factors to have an effect on the maximum drilling moment and RMS moment. Spectral analysis of the moment signal indicated that the drilling conditions produced no statistically significant effects in the amplitude of the spectral density at the cutting tooth frequency. Cutting speed was the only significant factor that contribute to the amplitude of the spectral density and the spindle frequency.

Conclusion

From the results of this research the following conclusions have been drawn:

- HSS/Cobalt drills produce holes with the greatest amount of maximum thrust force and RMS thrust force, while carbide drills produce holes with the least amount of maximum thrust force. The maximum thrust force produced by PCD drills are a little higher than that produced by carbide drills.
- Maximum thrust force and RMS thrust force increase with increasing feed rate and decreases with an increasing cutting speed for all drill tool materials tested.
- Of the three drill point angles of 90°, 118°, and 135° investigated, holes drilled with drill point angle of 118° produced the highest maximum thrust force, while drills with drill point angles of 90° produced the lowest amount of maximum thrust force. Drill point angle of 118° produced maximum thrust force slightly less than that produced by drills with 135° point angle.
- Drill point angle of 135° was found to produce the largest RMS thrust force followed by 118° point angle for Carbide and PCD material drills, while drills with 118° point angle produced the highest RMS thrust force when drilling with HSS/Cobalt drills.
- Drill material, feed rate and cutting speed, contribute approximately 77% of the variance measured in the maximum thrust force and 82% of the variance measured in the RMS thrust force.
- In order to minimize the maximum thrust and RMS thrust force generated when drilling AS4/PEEK composites, the drilling conditions of 90° point angle, carbide drill, feedrate of 50.8 mm/min (2 ipm) and spindle speed of 6000 rpm should be selected.

- The effect of cutting speed on the RMS moment and maximum moment is non-linear, with the greatest moment produced by holes drilled at 4000 rpm. Holes drilled at 2000 rpm produced the least moment, followed by 6000 rpm.
- Carbide drills produced holes with the least amount of RMS moment and maximum moment while HSS/Cobalt material drills produce the greatest RMS moment and maximum moment.
- Spindle speed and drill material contributed to approximately 50% of the variance measured in the RMS moment and 36% of the variance measured in the maximum moment.
- Spindle speed contributed 14% of the variance measured in the amplitude of the spectral density at the spindle frequency.
- The drilling conditions were found to have no significant effect on the magnitude of the spectral density at the tooth frequency.

II. EFFECTS OF DRILLING CONDITIONS, DRILL MATERIAL AND POINT ANGLE ON HOLE QUALITY AND DELAMINATION IN DRILLING ADVANCED COMPOSITE MATERIAL.

Abstract

A 1/3 fractional factorial design of experiment with 4 factors at 3 levels was used to study drilling of advanced multi-directional carbon fiber reinforced composites. Series of 1/4 inch holes were drilled in a 0.3 inch thick AS4/PEEK composite plate as per the 1/3 fractional factorial design of experiment. The plate was mounted on a specially designed fixture which allowed for the acquisition of thrust force, drilling torque (moment) from a Kistler 9272 four component dynamometer. The voltage proportional signals were separated via Kistler Type 5405A junction box and fed into separate Kistler Type 5010 dual mode amplifiers. The amplified signals were ran through low pass filters to reduce the unwanted noise from the drilling process. The digitized signals were save and transferred to a Pentium PC for further processing. Acoustic emission signals generated by the drilling process was picked up using Digital Wave B-1025 broadband sensors mounted on the AS4/PEEK composite plate and to the drilling fixture. The AE signals were passed through Digital Wave PA2040GA AE pre-amplifiers with 40 dB gain. The voltage proportional analog signals were fed into a Digital wave F4000 Fracture Wave Detector where the signals were filtered with band-pass filters, digitized at a sampling frequency of 6.25 MHz with record length of 2048 data points and saved directly to a Pentium PC. The Thrust force, drilling torque, and acoustic emission signals were acquired for the entire length of each hole drilling cycle. The design of experiment was conducted to determine the main effects and two factor interactions of: drill point angle, cutting speed, feed rate, and drill material on hole quality and

delamination in the drilling of AS4/PEEK composite plate. The two factor interactions of point angle and drill tool material, speed and drill tool material, as well as the main effect of drill point angle were found to have statistically significant effect on hole diameter. Drill tool material, point angle and drill tool material interaction, and the main effect of drill point angle were found to have a statistically significant effect on hole circularity. Drill tool material was found to be the most statistically significant factor and accounted for over 50% of the variation measured in the delamination of the exit hole side of the composite material.

Conclusion:

From the results of this research, the following conclusions have been drawn:

- PCD drills produce the most consistent size holes throughout the different ranges of cutting speeds and drill point angles.
- Increasing the drill point angle has a tendency to produce nominal size holes, while the effects of feed rate are found statistically insignificant.
- The drilling conditions of 118° point angle, PCD drill material, 50.8 mm/min (2ipm) and 2000 rpm should be selected to produce nominal size holes in drilling AS4/PEEK.
- Carbide tools were found to produce drilled holes with the least amount of circularity error through out all ranges of drill point angles tested followed by PCD drills, while HSS/Cobalt drills produced holes with the greatest amount of circularity.
- Drills with 118° drill point angles were found to produce the truest holes for all materials measured.
- The effects of feed rate and cutting speed on the circularity measurement were found to be statistically insignificant.
- The drilling conditions of 135° point angle, carbide drill with moderate cutting speeds and feed rates should be selected to produce hole with the least amount of circularity error in drilling AS4/PEEK.
- Drill material selection was found to have the most significant effect on exit hole delamination in drilling AS4/PEEK composites.
- Carbide drills were found to produce the least amount of exit delamination throughout all ranges of speed followed by PCD drills, while HSS/Cobalt drills were found to produce holes with the greatest amount of hole exit delamination.
- High feed rates produce the least amount of exit delamination when using drills with a drill point angle of 118° or 135°.

- The drilling conditions of 118° point angle, carbide drill material, 152 mm/min (6ipm) and 4000 rpm should be selected to minimize exit delamination in drilling AS4/PEEK.
- The higher the point angle the smaller the exit hole delamination for all drill tool materials. Point angle of 90° gave the highest delamination.

III. EFFECTS OF TOOL WEAR ON THRUST FORCE, TORQUE AND ACOUSTIC EMISSION IN DRILLING ADVANCED COMPOSITE MATERIAL

Abstract

Additional experimental investigation was conducted using the optimum drilling conditions identified in the previous research (previous two papers) to study drill wear and drill wear characteristics for the identified two best drill tool materials (carbide and PCD drills). The effects of drill wear on thrust force, cutting torque and acoustic emission for both carbide and PCD drills when drilling the advanced composite plate (AS4/PEEK) were investigated. SEM photographs of drilled specimen showing extent of delamination and internal cracks were taken and used to corroborate the analysis from sensor data and measured tool wear. Drilling conditions were selected to minimize exit hole delamination for carbide drill and a series of 80 holes were drilled for both carbide and PCD drills. The thrust force, drilling torque, acoustic emission were acquired for the entire length of drilling cycle for each hole. Carbide drill was found to wear more than PCD drill. Maximum and RMS thrust force were found to increase with tool wear. The maximum drilling torque was found to increase with drill wear for the first 45 holes drilled after which it began to decrease for both Carbide and PCD. No correlation was found between the RMS of drilling torque, the spectral density amplitude of the torque signal at tooth frequency with drill wear. The average acoustic emission (AE) energy as well as the total AE energy was found to decrease with tool wear.